

T.; Artlich, P.; Kaine, B.P.; Sykes, S.M.; Sadow, P.W.;  
D'Andrea, K.P.; Bowman, C.; Fujii, C.; Garland, S.A.;  
Mason, T.M.; Olsen, G.J.; Fraser, C.M.; Smith, H.O.; Woese,  
C.R.; Venter, J.C.

#journal Nature (1997) 390:364-370  
#title The complete genome sequence of the hyperthermophilic,  
sulfate-reducing archaeon *Archaeoglobus fulgidus*.

#cross-references GB:98049343  
#accession D69373

##status preliminary; nucleic acid sequence not shown;

##molecule\_type DNA

##residues 1-347 #label KLE

##cross-references GB:AF001036; GB:AF000782; NID:g2649610;  
TIGR:AF0988

SUMMARY #length 347 #molecular-weight 38112 #checksum 5546

Query Match 78.5%; Score 51; DB 2; Length 347;

Best Local Similarity 75.0%; Pred. No. 3.10e+00;

Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 273 PDAVYKL 280

QY 2 PDAVYKL 9

RESULT 3

ENTRY

TITLE semaphorin III - mouse (fragment)

ORGANISM #formal\_name Mus musculus #common\_name house mouse

DATE 26-Jul-1996 #sequence\_revision 28-Jul-1996 #text\_change

04-Sep-1998

ACCESSIONS I58169

REFERENCE I58169

#authors Messersmith, E.K.; Leonardo, E.D.; Shatz, C.J.;

Tessier-Lavigne, M.; Goodman, C.S.; Kolodkin, A.L.

#journal Neuron (1995) 14:949-959

#title Semaphorin III can function as a selective chemorepellent to

pattern sensory projections in the spinal cord.

#cross-references MUID:95267432

#accession I58169

##status preliminary; translated from GB/EMBL/DBJ

##molecule\_type mRNA

##residues 1-666 #label RES

##cross-references GB:L40484; NID:g703189; PID:g703190

GENETICS

#gene SemIII

CLASSIFICATION #superfamily semaphorin

SUMMARY #length 666 #checksum 9654

Query Match 76.9%; Score 50; DB 2; Length 666;

Best Local Similarity 44.4%; Pred. No. 5.03e+00;

Matches 4; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 43 HPDNIIFKL 51

QY 1 QPDDAVYKL 9

RESULT 4

ENTRY

TITLE semaphorin III precursor - human

ORGANISM #formal\_name Homo sapiens #common\_name man

DATE 06-Jan-1995 #sequence\_revision 06-Jan-1995 #text\_change

04-Sep-1998

ACCESSIONS D49423

REFERENCE A49423

#authors Kolodkin, A.L.; Matthes, D.J.; Goodman, C.S.

#journal Cell (1993) 75:1389-1399

#title The Semaphorin genes encode a family of transmembrane and

secreted growth cone guidance molecules.

#accession D49423

##status preliminary; nucleic acid sequence not shown

##molecule\_type mRNA

##residues 1-771 #label KOL

##cross-references GB:L26081; NID:g799328; PID:g436560

GENETICS

#gene GDB:SEMAL

##cross-references GDB:283448

CLASSIFICATION #superfamily semaphorin

SUMMARY #length 771 #molecular-weight 8889 #checksum 6249

Query Match 76.9%; Score 50; DB 2; Length 771;

Best Local Similarity 44.4%; Pred. No. 5.03e+00;

Matches 4; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 149 HPDNIIFKL 157

QY 1 QPDDAVYKL 9

RESULT 5

ENTRY

TITLE semaphorin D - mouse

ORGANISM #formal\_name Mus musculus #common\_name house mouse

DATE 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change

04-Sep-1998

ACCESSIONS I48747

REFERENCE I48747

#authors Puschel, A.W.; Adams, R.H.; Betz, H.

#journal Neuron (1995) 14:941-948

#title Murine semaphorin D/collapsin is a member of a diverse gene

family and creates domains inhibitory for axonal extension.

#cross-references MUID:95267431

#accession I48747

##status preliminary; translated from GB/EMBL/DBJ

##molecule\_type mRNA

##residues 1-772 #label RES

##cross-references EMBL:X85993; NID:g854329; PID:g854330

GENETICS

#gene semD

CLASSIFICATION #superfamily semaphorin

SUMMARY #length 772 #molecular-weight 88710 #checksum 1776

Query Match 76.9%; Score 50; DB 2; Length 772;

Best Local Similarity 44.4%; Pred. No. 5.03e+00;

Matches 4; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 149 HPDNIIFKL 157

QY 1 QPDDAVYKL 9

RESULT 6

ENTRY

TITLE anaphase spindle elongation protein ASEL - yeast

(Saccharomyces cerevisiae)

protein O2806; protein YOR088c

#formal\_name Saccharomyces cerevisiae

DATE 13-Jan-1996 #sequence\_revision 01-Mar-1996 #text\_change

06-Feb-1998

ACCESSIONS S59660; S66941

REFERENCE S59660

#authors Pellman, D.; Fink, G.R.

#submission submitted to the EMBL Data Library, January 1995

#description Yeast microtubule-associated proteins required for anaphase

spindle elongation.

#accession S59660

##molecule\_type DNA

##residues 1-885 #label PEL

##cross-references EMBL:U20235; NID:g972941; PID:g972942

REFERENCE S66929

#authors Bohn, C.; Bolotin-Fukuhara, M.; Daignan-Fornier, B.; Dang,

D.V.; Valens, M.

#submission submitted to the Protein Sequence Database, July 1996

#accession S66941

\*\*\*\*\*  
[M][P][E][L] (TM)  
\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Fri Apr 14 23:37:12 2000; MasPar time 3.31 Seconds  
Tabular output not generated. 108.824 Million cell updates/sec

Title: >US-08-452-843-7  
Description: (1-9) from US08452843.pep  
Perfect Score: 65  
Sequence: 1 QPDDAVYKL 9

Scoring table: PAM 150  
Gap 15

Searched: 122810 seqs, 40068593 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: p1r62  
1:p1r1 2:p1r2 3:p1r3 4:p1r4

Statistics: Mean 23.304; Variance 27.733; scale 0.840

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

				SUMMARIES			
				Query			
Result No.	Score	Match	Length	DB	ID	Description	Pred. No.
1	58	89.2	163	2	A54261	olfactory marker prot	8.70e-02
2	51	78.5	347	2	D59373	immunogenic protein (	3.10e+00
3	50	76.9	666	2	I38169	semaphorin III - mous	5.03e+00
4	50	76.9	771	2	D49423	semaphorin III precu	5.03e+00
5	50	76.9	772	2	I48747	semaphorin D - mous	5.03e+00
6	50	76.9	885	2	S59660	anaphase spindle elon	5.03e+00
7	50	76.9	1213	2	E69255	mannosyltransferase A	5.03e+00
8	49	75.4	773	1	Q3R8G	secretory component p	8.13e+00
9	49	75.4	832	2	S76815	hypothetical protein	8.13e+00
10	49	75.4	1021	2	S64506	protein kinase BUB1 (	8.13e+00
11	49	75.4	1591	2	A54146	invasion-inducing pro	8.13e+00
12	48	73.8	508	2	A33378	fasciclin III precurs	1.30e+01
13	48	73.8	560	2	S27387	interferon alpha rece	1.30e+01
14	47	72.3	162	2	A27450	olfactory marker prot	2.07e+01
15	47	72.3	163	2	B94261	olfactory marker prot	2.07e+01
16	47	72.3	492	2	S71245	glucose-6-phosphate 1	2.07e+01
17	47	72.3	509	2	S47553	cytochrome P450 Cyp4a	2.07e+01
18	47	72.3	563	2	T00338	L-lactate permease ho	2.07e+01
19	47	72.3	588	2	A70340	glucose-6-phosphate 1	2.07e+01
20	47	72.3	599	2	T00659	glucose-6-phosphate 1	2.07e+01
21	47	72.3	1215	2	S0904	hypothetical protein	2.07e+01
22	47	72.3	2211	1	EF05	coagulation factor V	2.07e+01
23	46	70.8	254	2	D70406	DMSO reductase chain	3.28e+01

24	46	70.8	308	1	PABY3	phosphoprotein phosph	3.28e+01
25	46	70.8	362	2	S17285	hypothetical protein	3.28e+01
26	46	70.8	387	2	S77268	carboxynorspermidine	3.28e+01
27	46	70.8	434	1	A35005	u-plasminogen activat	3.28e+01
28	46	70.8	515	2	A56686	glucose-6-phosphate 1	3.28e+01
29	46	70.8	515	1	S01233	glucose-6-phosphate 1	3.28e+01
30	46	70.8	515	1	DEHUG6	glucose-6-phosphate 1	3.28e+01
31	46	70.8	656	2	A56975	VI polysaccharide cap	3.28e+01
32	46	70.8	698	2	A47203	protein-glutamine gam	3.28e+01
33	46	70.8	787	2	H70374	NADH dehydrogenase I	3.28e+01
34	46	70.8	1392	2	T01908	hypothetical protein	3.28e+01
35	45	69.2	171	2	JH0246	phosphinothricin N-ac	5.14e+01
36	45	69.2	262	2	PC4159	ribosomal protein S4	5.14e+01
37	45	69.2	263	2	I48169	ribosomal protein S4	5.14e+01
38	45	69.2	289	2	A37209	thiosulfate sulfurtra	5.14e+01
39	45	69.2	320	1	A39479	homeotic protein mec-	5.14e+01
40	45	69.2	360	2	S52662	S-adenosylmethionine	5.14e+01
41	45	69.2	508	2	S77559	threonine dehydratase	5.14e+01
42	45	69.2	880	2	S03601	RNA-binding protein V	5.14e+01
43	45	69.2	881	1	P2XRUK	RNA-binding protein -	5.14e+01
44	45	69.2	881	1	P2XRSR	RNA-binding protein -	5.14e+01
45	45	69.2	1104	2	A36866	microbial collagenase	5.14e+01

ALIGNMENTS

RESULT 1

ENTRY A54261 #type complete  
TITLE Olfactory marker protein - human  
ORGANISM #formal\_name Homo sapiens #common\_name man  
DATE 09-Sep-1994 #sequence\_revision 09-Sep-1994 #text\_change 09-Sep-1994

ACCESSIONS A54261  
REFERENCE A54261  
#authors F.L.

#journal Genomics (1994) 20:452-462  
#title Human and rodent OMP genes: conservation of structural and regulatory motifs and cellular localization.

#accession A54261  
#status preliminary  
#molecule\_type DNA  
#residues 1-163 #label BUI  
#cross-references GB:U01212  
SUMMARY #length 163 #molecular-weight 19064 #checksum 8002

Query Match 89.2%; Score 58; DB 2; Length 163;  
Best Local Similarity 55.6%; Pred. No. 8.70e-02;  
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db	47	QPESVYRL	55	
Qy	1	QPDDAVYKL	9	

RESULT 2

ENTRY D69373 #type complete  
TITLE immunogenic protein (bcspl-3) homolog - Archaeoglobus fulgidus

ORGANISM #formal\_name Archaeoglobus fulgidus  
DATE 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 05-Jun-1998

ACCESSIONS D69373  
REFERENCE A69250

#authors Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson, R.J.; Gwinn, M.; Hickey, E.K.; Peterson, J.D.; Richardson, D.L.; Kerlavage, A.R.; Graham, D.E.; Kyrpides, N.C.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.; Dougherty, B.A.; McKenny, K.; Adams, M.D.; Loftus, B.; Peterson, S.; Reich, C.I.; McNeill, L.K.; Badger, J.H.; Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.; Utterback, T.; Cotton, M.D.; Spriggs,



PD 18-MAR-1993.  
 PF 27-AUG-1992; U07289.  
 PR 28-AUG-1991; US-750913.  
 PR 06-JAN-1992; US-817912.  
 PA (UYPE-) UNIV PENNSYLVANIA.  
 PA (WIST-) WISTAR INST.  
 PI Weiner DB, Williams WV;  
 DR WPI; 93-100655/12.  
 DR N-PSDB; Q37082.  
 PT T-cell receptor based treatment of rheumatoid arthritis - comprises  
 PT administration of antibodies to T-cell receptor variable regions  
 PS Disclosure: Page 24; 110pp; English.  
 CC The sequence is that of human rheumatoid synovial T cell receptor alpha  
 CC chain clone alpha17.4 from patient #4. It may be used, as part of  
 CC a method of treating rheumatoid arthritis, to raise antibodies which  
 CC can be administered to treat the arthritis. This therapeutic  
 CC approach to treatment of rheumatoid arthritis involves deletion of  
 CC only those T cells involved in the autoimmune response. Since these  
 CC comprise only a small portion of the total T cell repertoire,  
 CC eliminating these T cells should not result in significant  
 CC generalised immunosuppression. It may also be used in immunisation  
 CC to prevent the occurrence of rheumatoid arthritis.  
 SQ Sequence 117 AA;

Query Match 56.2%; Score 43; DB 1; Length 117;  
 Best Local Similarity 75.0%; Pred. No. 3.99e+02;  
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 102 PDPAYVOL 109  
 QY 2 PDDAVYKL 9

Search completed: Fri Apr 14 23:36:54 2000  
 Job time : 42 secs.

SQ Sequence 113 AA;

Query Match 66.2%; Score 43; DB 1; Length 113;  
Best Local Similarity 75.0%; Pred. No. 3.99e+02;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 5 PDPVYQL 12  
II III I  
QY 2 PDDAVYKL 9

RESULT 12

ID R33269 standard; Protein; 114 AA.  
AC R33269;  
DT 16-JUL-1993 (first entry)  
DE T cell receptor alpha chain clone alpha1.1/2.  
KW Rheumatoid arthritis; synovial; therapy; therapeutic;  
KW autoimmunity response; variable region; mammal; immunisation.  
OS Homo sapiens.  
PN WO9304695-A.  
PD 18-MAR-1993.  
PF 27-AUG-1992; U07289.  
PR 28-AUG-1991; US-750913.  
PR 06-JAN-1992; US-817912.  
PA (UYPE-) UNIV PENNSYLVANIA.  
PA (WIST-) WISTAR INST.  
PI Weiner DB, Williams WV;  
DR WPI: 93-100655/12.  
DR N-PSDB; Q37076.  
PT T-cell receptor based treatment of rheumatoid arthritis - comprises  
PT administration of antibodies to T-cell receptor variable regions  
PS Disclosure; Page 18; 110pp; English.  
CC The sequence is that of human rheumatoid synovial T cell receptor alpha  
CC chain clone alpha1.1/2 from patient #1. It may be used, as part of  
CC a method of treating rheumatoid arthritis, to raise antibodies which  
CC can be administered to treat the arthritis. This therapeutic  
CC approach to treatment of rheumatoid arthritis involves deletion of  
CC only those T cells involved in the autoimmune response. Since these  
CC comprise only a small portion of the total T cell repertoire,  
CC eliminating these T cells should not result in significant  
CC generalised immunosuppression. It may also be used in immunisation  
CC to prevent the occurrence of rheumatoid arthritis.  
SQ Sequence 114 AA;

Query Match 66.2%; Score 43; DB 1; Length 114;  
Best Local Similarity 75.0%; Pred. No. 3.99e+02;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 99 PDPVYQL 106  
II III I  
QY 2 PDDAVYKL 9

RESULT 14

ID R33274 standard; Protein; 117 AA.  
AC R33274;  
DT 16-JUL-1993 (first entry)  
DE T cell receptor alpha chain clone alpha1.7.3.  
KW Rheumatoid arthritis; synovial; therapy; therapeutic;  
KW autoimmunity response; variable region; mammal; immunisation.  
OS Homo sapiens.  
PN WO9304695-A.  
PD 18-MAR-1993.  
PF 27-AUG-1992; U07289.  
PR 28-AUG-1991; US-750913.  
PR 06-JAN-1992; US-817912.  
PA (UYPE-) UNIV PENNSYLVANIA.  
PA (WIST-) WISTAR INST.  
PI Weiner DB, Williams WV;  
DR WPI: 93-100655/12.  
DR N-PSDB; Q37081.  
PT T-cell receptor based treatment of rheumatoid arthritis - comprises  
PT administration of antibodies to T-cell receptor variable regions  
PS Disclosure; Page 23; 110pp; English.  
CC The sequence is that of human rheumatoid synovial T cell receptor alpha  
CC chain clone alpha1.7.3 from patient #4. It may be used, as part of  
CC a method of treating rheumatoid arthritis, to raise antibodies which  
CC can be administered to treat the arthritis. This therapeutic  
CC approach to treatment of rheumatoid arthritis involves deletion of  
CC only those T cells involved in the autoimmune response. Since these  
CC comprise only a small portion of the total T cell repertoire,  
CC eliminating these T cells should not result in significant  
CC generalised immunosuppression. It may also be used in immunisation  
CC to prevent the occurrence of rheumatoid arthritis.  
SQ Sequence 117 AA;

Query Match 66.2%; Score 43; DB 1; Length 117;  
Best Local Similarity 75.0%; Pred. No. 3.99e+02;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 102 PDPVYQL 109  
II III I  
QY 2 PDDAVYKL 9

RESULT 15

ID R33275 standard; Protein; 117 AA.  
AC R33275;  
DT 16-JUL-1993 (first entry)  
DE T cell receptor alpha chain clone alpha1.7.4.  
KW Rheumatoid arthritis; synovial; therapy; therapeutic;  
KW autoimmunity response; variable region; mammal; immunisation.  
OS Homo sapiens.  
PN WO9304695-A.

Db 4 PDPAVYQL 11

OS Homo sapiens.  
PN W09743411-AL.  
PD 20-NOV-1997.  
PF 09-MAY-1997; J01565.  
PR 29-MAY-1996; JP-135572.  
PR 10-MAY-1996; JP-116101.  
PA (KIRI ) KIRIN BEER KK.  
PI Honma N, Mikayama T, Yuyama N;  
DR WFI; 98-008880/01.  
PT Immunosuppressant peptide containing T-cell receptor alpha-chain  
PT sequence - are not antigen-specific and do not induce antibody  
production  
PS Claim 3; Page 36; 63pp; Japanese.  
CC The present sequence represents human T-cell receptor alpha-chain  
CC constant region. The protein is an immunosuppressant which is not  
CC antigen-specific and suppresses both humoral, and cell-mediated  
CC reactions. It can be used for treatment and/or prevention of del  
CC hypersensitivity reactions, allergies and autoimmune reactions,  
CC inhibition of transplant rejection. The protein does not induce  
CC formation of antibodies against them to any significant extent.

PS Claim 10; Pages 99-102; 152pp; English.  
 CC The present sequence is the rabbit poly-immunoglobulin (Ig)  
 CC receptor, a portion of which corresp. to residues 1-627, pref.  
 CC 1-606, or esp. residues 21-43, 1-118, 119-223, 224-332, 333-441,  
 CC 442-552, 553-606 or 553-627 comprises a protection protein (PP).  
 CC The Ig of the invention comprises a PP as above in association with  
 CC an Ig derived heavy chain, having at least a portion of an antigen  
 CC (Ag) binding domain. The PP protects the Ig in harsh mucosal, e.g.  
 CC gastrointestinal, environments, therefore enhancing its  
 CC effectiveness in passively immunising animals against mucosal  
 CC pathogens. The Ag binding domain is specifically derived from the  
 CC Guy's 13 antibody, and the Ig can be used to prevent dental caries  
 CC by binding, e.g. Streptococcus mutans serotypes c, e and f, or  
 CC S. sorbinus serotypes d and g.  
 SQ Sequence 773 AA;

Query Match 75.4%; Score 49; DB 1; Length 773;  
 Best Local Similarity 75.0%; Pred. No. 8.52e+01;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 137 EPDDVYK 144  
 :||| |||  
 QY 1 QPDDAVYK 8

RESULT 5  
 ID R05881 standard; protein; 198 AA.  
 AC R05881;  
 DT 27-NOV-1990 (first entry)  
 DE Sequence encoded by clone 61.  
 KW Interleukin-2; IL-2; cancer; ds.  
 OS Homo sapiens.  
 PN US4939093-A.  
 PD 03-JUL-1990.  
 PF 23-AUG-1988; 236296.  
 PR 28-SEP-1982; US-426059.  
 PR 13-JAN-1983; US-457594.  
 PR 02-FEB-1987; US-009999.  
 PR 23-AUG-1988; US-236296.  
 PA (CETU ) CETUS CORP.  
 PI MCGROGAN MP, KAWASAKI ES, DOYLE MV, MARK DF;  
 DR WPI; 90-224018/29.  
 DR N-PSDB; Q05237.  
 PT Messenger RNA expressing interleukin 2 in X.laevs oocyte -  
 PT isolated by hybridisation with new recombinant DNA, also useful  
 PT for expression in bacterial hosts.  
 PS Disclosure; P; English.  
 CC Clone may be used to produce IL-2 in a X.laevs oocyte translation  
 CC system. IL-2 is useful in diagnosis and treatment of cancer,  
 CC infections and immune diseases.  
 SQ Sequence 198 AA;

Query Match 69.2%; Score 45; DB 1; Length 198;  
 Best Local Similarity 62.5%; Pred. No. 2.40e+02;  
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 51 PEEAKYKL 58  
 :||| |||  
 QY 2 PDDAVYKL 9

RESULT 6  
 ID R38151 standard; Protein; 580 AA.  
 AC R38151;  
 DT 13-OCT-1993 (first entry)  
 DE Acetobacter diguanylate cyclase DGC1.  
 KW Cyclic diguanylate; diguanylate phosphodiesterase;  
 KW diguanylate cyclase; cellulose production; cdi operon.  
 OS Acetobacter xylinum.  
 FH Key Location/Qualifiers  
 FT misc\_difference 238  
 FT /note- "deduced from CCG (Pro) codon"  
 FT misc\_difference 353

FT W09311244-A.  
 PN 10-JUN-1993.  
 PD 14-OCT-1992; U08756.  
 PF 29-NOV-1991; US-800218.  
 PR (WEYE ) WEYERHAEUSER CO.  
 PA Ben-Bassat A, Ben-Ziman M, Calhoon RD, Gelfand DH;  
 PI Tal R, Wong HC;  
 PR WPI; 93-197062/24.  
 DR N-PSDB; Q43660.  
 PT Polynucleotide sequence from Acetobacter cdi operon - encodes  
 PT cyclic di:guanosine mono:phosphate degradation enzymes e.g.  
 PT 3-phosphodiesterase isozyme  
 PS Claim 5; Page 77-79; 98pp; English.  
 CC The amino acid sequence of protein DGC1 was deduced from the third  
 CC open reading frame of the cdi operon. The protein has diguanylate  
 CC cyclase activity, i.e. it enzymatically converts two molecules of  
 CC GTP to bis-(3',5')-cyclic diguanylic acid.  
 CC See also R38149-R38150 and R38152.  
 SQ Sequence 580 AA;

Query Match 69.2%; Score 45; DB 1; Length 580;  
 Best Local Similarity 55.6%; Pred. No. 2.40e+02;  
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 213 HPEDAVYK 221  
 :||| |||  
 QY 1 QPDDAVYK 9

RESULT 7  
 ID R06723 standard; protein; 653 AA.  
 AC R06723;  
 DT 18-JAN-1991 (first entry)  
 DE Achromobacter protease I.  
 KW T-API; enzyme prodn.; peptide mapping; peptide synthesis.  
 OS Achromobacter lyticus.  
 PN EP-387646-A.  
 PD 19-SEP-1990.  
 PF 03-MAR-1990; 104163.  
 PR 14-MAR-1989; JP-059726.  
 PA (WAKP ) WAKO PURE CHEM IND KK.  
 PI Sakiyama F, Nakata A;  
 DR WPI; 90-283902/38.  
 DR N-PSDB; Q05926.  
 PT Novel DNA encoding Achromobacter protease I - for recombinant  
 PT prodn. of enzyme, and for fragmentation of protein(s) and  
 PT peptide, for peptide mapping and synthesis of lys-x-cpds.  
 PS Disclosure; fig 1; 20pp; English.  
 CC This Achromobacter protease I or an analogue (T-API) specific-  
 CC ally cleaves the peptide bonds (-Lys-X-) on the side of the  
 CC carboxyl gps. of lysine residues in proteins and peptides  
 CC all Lys-x bonds are cleaved incl. the Lys-Pro bond. T-APIs  
 CC are therefore useful for fragmenting proteins or peptides  
 CC for primary structural analysis, prodn. of peptide maps or  
 CC the synthesis of -Lys-X- cpds.  
 SQ Sequence 653 AA;

Query Match 67.7%; Score 44; DB 1; Length 653;  
 Best Local Similarity 62.5%; Pred. No. 3.10e+02;  
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 614 PDGTVYKL 621  
 :||| |||  
 QY 2 PDDAVYKL 9

RESULT 8  
 ID R33271 standard; Protein; 108 AA.  
 AC R33271;  
 DT 16-JUL-1993 (first entry)  
 DE T cell receptor alpha chain clone alpha1.4.  
 KW Rheumatoid arthritis; synovial; therapy; therapeutic;

DE Human semaphorin III protein.  
KW Semaphorin; grasshopper; human; vaccinia virus; Drosophila; Tribolium;  
KW variola major virus; smallpox; semaphorin receptor binding activity;  
KW modulation; nerve cell growth; immune response; viral pathogenesis;  
KW neurological disease; neuro-regeneration; oncological infection.  
OS Homo sapiens.  
PN W09507706-A.  
PD 23-MAR-1995.  
PF 13-SEP-1994; U10151.  
PR 13-SEP-1993; US-121713.  
PA (REGC ) UNIV CALIFORNIA.  
PI Bentley DR, Goodman CS, Kolodkin AL, Matthes D;  
PI O'Connor T;  
PI WPI; 95-131177/17.  
DR N-PSDB; Q87442.  
PT New class of semaphorin peptide(s) and polypeptide(s) - are  
PT potent modulators of nerve cell growth and regeneration  
PS Example 2; Page 60-63; 101pp; English.  
CC The sequence of the human semaphorin III protein. The proteins  
CC encoded by the grasshopper semaphorin I (Q87441), human semaphorin III,  
CC vaccinia virus semaphorin IV (Q87443), Drosophila semaphorin I and II  
CC (Q87444-5), Tribolium semaphorin I (Q87446) or variola major (smallpox)  
CC virus semaphorin IV (Q87447) genes were used to generate a series of  
CC peptides (R70370-R70418), which retain semaphorin receptor binding  
CC activity. The semaphorin derived or semaphorin receptor derived peptides  
CC are potent modulators of nerve cell growth, immune responsiveness and  
CC viral pathogenesis. They can be used in diagnosis and treatment of  
CC neurological disease and neuro-regeneration, immune modulation and  
CC diagnosis and treatment of viral and oncological infection and diseases.  
SQ Sequence 771 AA;

Query Match 76.9%; Score 50; DB 1; Length 771;  
Best Local Similarity 44.4%; Pred. No. 6.54e+01;  
Matches 4; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 149 HPEDNIFKL 157

QY 1 QPDDAVIKL 9

RESULT 3  
ID R14670 standard; Protein; 584 AA.  
AC R14670;  
DE 30-JAN-1992 (first entry)  
DE Truncated poly Ig-receptor encoded by allele no. 1.  
KW Rabbit; insemination; pregnancy.  
OS Oryctolagus cuniculus.  
FH Key Location/Qualifiers  
FT peptide 1..18  
FT domain 10..118  
FT /label= signal sequence  
FT /number= I  
FT /note="poly-Ig binding"  
FT domain 119..223  
FT /number= II  
FT domain 224..332  
FT /number= III  
FT domain 333..441  
FT /number= IV  
FT domain 442..552  
FT /number= V  
FT domain 553..584  
FT /number= IV  
FT /note="incomplete"

PN W09116061-A.  
PD 31-OCT-1991.  
PF 16-APR-1991; U02604.  
PR 16-APR-1990; US-510161.  
PA (HARD ) HARVARD COLLEGE.  
PA (SURE-) INST SUISSE RECH EXPR C.  
PI Kraehenbuhl JP, Weltzin RA, Neutra MR;  
DR WPI; 91-339549/46.  
DR N-PSDB; Q14498.

PT Stabilised poly-Ig complex contg. portion of poly-Ig receptor -  
PT useful in protection against pathogens or against pregnancy  
PS Disclosure; Fig 7; 51 pp; English.  
CC The sequence was deduced from a cDNA clone of allele no. 1 and  
CC is a truncated poly-Ig receptor. The native gene (Mostov et al)  
CC intra-cellular domains. The recombinant protein produced by  
CC expression of the sequence is used as a stabiliser protein with a  
CC poly-Ig specific for a selected antigen or family of antigens. The  
CC compsn. can be administered directly to the mucosal surfaces of a  
CC mammal to protect against a pathogen or against insemination. It  
CC protects against allergens that contact the respiratory or digestive  
CC mucosal surfaces and protects against pregnancy by cross-linking  
CC sperm in the vagina.  
CC See also R14671.  
SQ Sequence 584 AA;

Query Match 75.4%; Score 49; DB 1; Length 584;  
Best Local Similarity 75.0%; Pred. No. 8.52e+01;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 137 EPDDVYVK 144

QY 1 QPDDAVYK 8

RESULT 4  
ID W03177 standard; Protein; 773 AA.  
AC W03177;  
DE 24-FEB-1997 (first entry)  
DE Rabbit poly-immunoglobulin receptor.  
KW Rabbit; immunoglobulin; receptor; protection protein; mutants;  
KW heavy chain; antigen binding domain; protection; pathogen;  
KW mucosal; environment; gastrointestinal; passive; immunisation;  
KW Guy's 13 antibody; prevention; dental caries; Streptococcus;  
KW poly; sorbinus.  
OS Oryctolagus cuniculus.  
FH Key Location/Qualifiers  
FT region 21..43  
FT /note="immunoglobulin binding residues of domain I"  
FT domain 1..118  
FT /label= domain\_I  
FT domain 119..223  
FT /label= domain\_II  
FT domain 224..332  
FT /label= domain\_III  
FT domain 333..441  
FT /label= domain\_IV  
FT domain 442..552  
FT /label= domain\_V  
FT region 553..606  
FT /note="external portions of domain VI"  
FT region 553..627  
FT /note="external portions of domain VI"  
FT region 630..652  
FT /label= transmembrane\_segment  
FT region 653..755  
FT /label= intracellular\_portion

PN W09621012-A1.

PD 11-JUL-1996.  
PF 27-DEC-1995; U16889.  
PR 30-DEC-1994; US-367395.  
PR 04-MAY-1995; US-434000.  
PA (PLAN-) PLANT BIOTECHNOLOGY INC.  
PA (UNME-) UNITED MEDICAL & DENTAL SCHOOLS GUYS.  
PA (PLAN-) PLANET BIOTECHNOLOGY INC.  
PI Hiatt AC, Lehner T, Ma JKC;  
DR WPI; 96-333987/33.  
DR N-PSDB; T31287.  
PT Immunoglobulin and protection protein complex and its prodn. in  
PT plants - useful for passive immunisation against mucosal antigens,  
PT esp. against S. mutans and S. sorbinus to prevent dental caries

\*\*\*\*\*  
W P S R E L  
\*\*\*\*\*  
(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Fri Apr 14 23:36:12 2000; MasPar time 6.26 Seconds  
Tabular output not generated. 34.062 Million cell updates/sec

Title: >US-08-452-843-7  
Description: (1-9) from US08452843.pap  
Perfect Score: 65  
Sequence: 1 QPDDAVYKL 9

Scoring table: PAM 150  
Gap 15

Searched: 188963 seqs, 23686106 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: a-geneseq36  
l:geneseqp

Statistics: Mean 16.764; Variance 46.752; scale 0.359

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	65	100.0	9	1 R89368	Cw4 consensus peptide	1.04e+00
2	50	76.9	771	1 R71380	Human semaphorin III p	6.54e+01
3	49	75.4	584	1 R44670	Truncated poly Ig-rece	8.52e+01
4	49	75.4	773	1 W03177	Rabbit poly-immunoglob	8.52e+01
5	45	69.2	198	1 R05881	Sequence encoded by cl	2.40e+02
6	45	69.2	580	1 R38151	Acetobacter diguanylat	2.40e+02
7	44	67.7	653	1 R06723	Achromobacter protease	3.10e+02
8	43	66.2	108	1 R33271	T cell receptor alpha	3.99e+02
9	43	66.2	112	1 R36112	Human T-cell receptor	3.99e+02
10	43	66.2	112	1 R33270	T cell receptor alpha	3.99e+02
11	43	66.2	113	1 W36108	Human T-cell receptor	3.99e+02
12	43	66.2	114	1 R33269	T cell receptor alpha	3.99e+02
13	43	66.2	114	1 R33276	T cell receptor alpha	3.99e+02
14	43	66.2	117	1 R33274	T cell receptor alpha	3.99e+02
15	43	66.2	117	1 R33275	T cell receptor alpha	3.99e+02
16	43	66.2	117	1 R33272	T cell receptor alpha	3.99e+02
17	43	66.2	117	1 R33273	T cell receptor alpha	3.99e+02
18	43	66.2	125	1 W98876	H. pylori GHPO 1749 pr	3.99e+02
19	43	66.2	136	1 R33277	T cell receptor alpha	3.99e+02
20	43	66.2	169	1 V11037	H. pylori ORF hp3p2111	3.99e+02
21	43	66.2	267	1 W47588	T-cell receptor alpha-	3.99e+02
22	43	66.2	277	1 P60065	Sequence of a polypept	3.99e+02
23	43	66.2	306	1 W01731	T. gondii antigen part	3.99e+02

24	43	66.2	306	1 R12346	Toxoplasma gondii prot	3.99e+02
25	43	66.2	368	1 W98043	Lactobacillus delbruec	3.99e+02
26	43	66.2	452	1 R12353	Toxoplasma gondii p88	3.99e+02
27	43	66.2	452	1 W01732	T. gondii antigen p88.	3.99e+02
28	43	66.2	593	1 R28349	Bacillus caldotenax DN	3.99e+02
29	43	66.2	685	1 W88432	Disease associated pro	3.99e+02
30	43	66.2	685	1 Y00915	Human serum inducible	3.99e+02
31	43	66.2	876	1 W35905	Bacillus stearothermop	3.99e+02
32	43	66.2	877	1 W2846	Bacillus caldotenax DN	3.99e+02
33	43	66.2	877	1 R28348	Bacillus caldotenax DN	3.99e+02
34	43	66.2	877	1 W22847	Bacillus caldotenax DN	3.99e+02
35	43	66.2	877	1 W22845	Bacillus caldotenax DN	3.99e+02
36	43	66.2	1074	1 R24102	Marek's disease virus	3.99e+02
37	43	66.2	2386	1 W13153	S. pombe Rad3 polypept	3.99e+02
38	42	64.6	109	1 W36107	Mouse T-cell receptor	5.12e+02
39	42	64.6	215	1 R77287	T-cell receptor alpha	5.12e+02
40	42	64.6	273	1 Y05405	Killer T-cell receptor	5.12e+02
41	42	64.6	274	1 Y05404	Killer T-cell receptor	5.12e+02
42	42	64.6	308	1 R45431	Diabetogene rad: A typ	5.12e+02
43	42	64.6	518	1 W61387	Schizophyllum commune	5.12e+02
44	42	64.6	521	1 W34477	RCH1-related protein.	5.12e+02
45	42	64.6	720	1 W98136	Human transglutaminase	5.12e+02

ALIGNMENTS

RESULT 1  
ID R89368 standard; peptide; 9 AA.  
AC R89368; 1996 (first entry)  
DT 18-SEP-1996  
DE Cw4 consensus peptide derived immunogenic peptide.  
KW Immunogenic peptide; supermotif; HLA molecule; CTL response;  
KW therapeutic; diagnostic; cancer; viral infection; hepatitis B;  
KW hepatitis C.  
OS Synthetic C.  
PN W05603140-A1.  
PD 08-FEB-1996.  
PF 21-JUL-1995; U09234.  
PR 21-JUL-1994; US-278634.  
PR 23-NOV-1994; US-344824.  
PR 30-MAY-1995; US-452843.  
PA (CYTE-) CYTEL CORP.  
PI Sette A, Sidney J;  
DR WPI; 96-116784/12.  
PT Compn. comprising immunogenic peptide with supermotif allowing more than one HLA mol. to bind - used to induce CTL response in patient and for in vivo and ex vivo therapeutic and diagnostic applications  
PS Claim 2; Page 26; 32pp; English.  
CC The sequences given in R89362-82 are immunogenic peptides which were use in the composition of the invention. The composition comprises an immunogenic peptide of 9-10 residues with a supermotif which allows binding of more than one HLA molecule. It pref. comprises two conserved residues, a first at the 2nd position from the N-terminal is Pro, and a 2nd at the C-terminal is Met. These peptides are used to induce a CTL response in a patient. They are also useful in compositions for in vivo and ex vivo therapeutic and diagnostic applications, e.g the treatment of cancer and viral infections, e.g. hepatitis B and C.  
SQ Sequence 9 AA;

Query Match 100.0%; Score 65; DB 1; Length 9;  
Best Local Similarity 100.0%; Pred. No. 1.04e+00;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Dd 1 QPDDAVYKL 9  
QY 1 QPDDAVYKL 9

RESULT 2  
ID R71380 standard; Protein; 771 AA.  
AC R71380;  
DT 21-NOV-1995 (first entry)

THIS PAGE BLANK (USPTO)

```

#accession A46345
#molecule_type DNA
#residues 1-395 #label BAM
#cross-references GB:M55567; NID:g215743; PID:g215744
#accession C46345
#molecule_type protein
#residues 2-10 #label BA2

GENETICS
#gene III
CLASSIFICATION #superfamily phage PRD1 gene III protein
KEYWORDS capsid protein
SUMMARY #length 395 #molecular-weight 43447 #checksum 2757

Query Match 100.0%; Score 19; DB 1; Length 395;
Best Local Similarity 25.0%; Pred.No. 1.53e+02;
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 138 PIKYGDM 145
QY 2 PXXXXXX 9

```

Search completed: Sat Apr 15 02:18:07 2000  
Job time : 20 secs.



```

COMMENT  This enzyme is involved in the Calvin cycle.
GENETICS
#gene      cbbFp; cxfF
#genome    plasmid
CLASSIFICATION
#superfamily fructose-bisphosphatase
KEYWORDS   gluconeogenesis; phosphoric monoester hydrolase
SUMMARY    #length 364 #molecular-weight 39777 #checksum 8838

Query Match      100.0%; Score 19; DB 2; Length 364;
Best Local Similarity 25.0%; Pred. No. 1.53e+02;
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 243 PRGKDFNM 250
|
QY 2 PXXXXXXM 9

RESULT 12
ENTRY S57273 #type complete
TITLE lignin peroxidase (EC 1.11.1.1-) LP7 precursor - white-rot
ORGANISM fungus
#formal_name Trametes versicolor #common_name white-rot
DATE 28-Oct-1995 #sequence_revision 03-Nov-1995 #text_change
S57273
#accessions
REFERENCE Johansson, T.; Nyman, P.O.
#authors Blochm. Biophys. Acta (1995) 1263:71-74
#journal The gene from the white-rot fungus Trametes versicolor
#title encoding the lignin peroxidase isozyme LP7.
#cross-references MUID:95959206
#accession S57273
#status preliminary
#molecule_type DNA
#residues 1-368 #label JOH
#cross-references EMBL:Z30667; NID:9495282; PID:9495283
GENETICS
#introns 21/1; 90/2; 135/2; 317/3; 342/3; 364/2
CLASSIFICATION
#superfamily lignin peroxidase
KEYWORDS extracellular protein; glycoprotein; heme; oxidoreductase
SUMMARY #length 368 #molecular-weight 39319 #checksum 7564

Query Match      100.0%; Score 19; DB 2; Length 368;
Best Local Similarity 25.0%; Pred. No. 1.53e+02;
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 215 PTIPGPM 222
|
QY 2 PXXXXXXM 9

RESULT 13
ENTRY B43673 #type complete
TITLE Chloromuconate cycloisomerase (EC 5.5.1.7) II - Pseudomonas
ORGANISM #formal_name Pseudomonas sp.
DATE 03-Mar-1993 #sequence_revision 03-Mar-1993 #text_change
17-Mar-1999
#accessions B43673
REFERENCE van der Meer, J.R.; Eggen, R.I.L.; Zehnder, A.J.B.; de Vos,
W.M.
#journal J. Bacteriol. (1991) 173:2425-2434
#title Sequence analysis of the Pseudomonas sp. strain P51 tcb gene
cluster, which encodes metabolism of chlorinated catechols:
evidence for specialization of catechol 1,2-dioxygenases
for chlorinated substrates.
#cross-references MUID:91193197
#accession B43673
#status preliminary
#molecule_type DNA
#residues 1-370 #label VAN

##cross-references GB:M57629; NID:gi151575; PID:gi151578
CLASSIFICATION #superfamily muconate cycloisomerase
KEYWORDS intramolecular lyase; isomerase
SUMMARY #length 370 #molecular-weight 39487 #checksum 3413

Query Match      100.0%; Score 19; DB 2; Length 370;
Best Local Similarity 25.0%; Pred. No. 1.53e+02;
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 14 PTRRPLQM 21
|
QY 2 PXXXXXXM 9

RESULT 14
ENTRY KIBSGM #type complete
TITLE phosphoglycerate kinase (EC 2.7.2.3) - Bacillus megaterium
ORGANISM #formal_name Bacillus megaterium
DATE 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change
05-Sep-1997
#accessions S13125; JQ1954
REFERENCE Schlaepfer, B.S.; Branlant, C.; Branlant, G.; Zuber, H.
#authors Nucleic Acids Res. (1990) 18:6423
#journal Nucleotide sequence of the phosphoglycerate kinase gene from
#title Bacillus megaterium.
#cross-references MUID:91057129
#accession S13125
#molecule_type DNA
#residues 1-394 #label SCH
#cross-references EMBL:X54519; NID:g39642; PID:g39643
JQ1952
#authors Schlaepfer, B.S.; Zuber, H.
#journal Gene (1992) 122:53-62
#title Cloning and sequencing of the genes encoding
glyceraldehyde-3-phosphate dehydrogenase, phosphoglycerate
kinase and triosephosphate isomerase (gap operon) from
mesophilic Bacillus megaterium: comparison with
corresponding sequences from thermophilic Bacillus
stearothermophilus.
#accession JQ1954
#molecule_type DNA
#residues 1-394 #label SC2
#experimental_source strain DSM319
GENETICS
#gene pgk
CLASSIFICATION #superfamily phosphoglycerate kinase
KEYWORDS ATP; gluconeogenesis; glycolysis; phosphotransferase
FEATURE #binding_site ATP (Lys, Glu) #status predicted
SUMMARY #length 394 #molecular-weight 42457 #checksum 2685

Query Match      100.0%; Score 19; DB 1; Length 394;
Best Local Similarity 25.0%; Pred. No. 1.53e+02;
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 164 PAVAGFLM 171
|
QY 2 PXXXXXXM 9

RESULT 15
ENTRY A46345 #type complete
TITLE gene III protein - phage PRD1
ALTERNATE_NAMES major capsid protein III
ORGANISM #formal_name phage PRD1
DATE 17-Feb-1994 #sequence_revision 17-Feb-1994 #text_change
05-Sep-1997
#accessions A46345; C46345
REFERENCE A46345
#authors Bamford, J.K.H.; Bamford, D.H.
#journal Virology (1990) 177:445-451
#title Capsomer proteins of bacteriophage PRD1, a bacterial virus

```

```

CLASSIFICATION #superfamily regulatory protein lysR
KEYWORDS DNA binding; modulation; transcription regulation
FEATURE
23-48 #region regulatory protein lysR motif
SUMMARY #length 314 #molecular-weight 35581 #checksum 2786

Query Match 100.0%; Score 19; DB 1; Length 314;
Best Local Similarity 25.0%; Pred. No. 1.53e+02;
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 206 PSLEEFM 213
QY 2 PXXXXXXM 9

RESULT 9 S68675 #type complete
ENTRY mitogen-activated protein kinase (EC 2.7.1.1) - fission yeast
TITLE (Schizosaccharomyces pombe)
ALTERNATE_NAMES MAP kinase; protein SPAC24B11.06c
ORGANISM #formal name Schizosaccharomyces pombe
DATE 25-Feb-1998 #sequence_revision 13-Mar-1998 #text_change
16-Dec-1998
ACCESSIONS S68675; S57930; S62551; S68433
REFERENCE #authors Kato Jr., T.; Okazaki, K.; Murakami, H.; Stettler, S.;
Fantes, P.A.; Okayama, H.
#journal FEBS Lett. (1996) 378:207-212
#title Stress signal, mediated by a Hog1-like MAP kinase, controls
sexual development in fission yeast.
#accession S68675
#molecule_type DNA
#residues 1-349 #label KAT
REFERENCE S57930
#authors Millar, J.B.A.; Buck, V.; Wilkinson, M.G.
#submission submitted to the EMBL Data Library, June 1995
#description Pyp1 and Pyp2 PTases dephosphorylate an osmosensing MAP
kinase controlling cell size at division in fission yeast.
#accession S57930
#molecule_type DNA
#residues 1-349 #label MIL
#cross-references EMBL:X89262; NID:g897809; PID:g897810
REFERENCE S62546
#authors Odell, C.; Churcher, C.M.
#submission submitted to the EMBL Data Library, November 1995
#accession S62551
#status preliminary
#molecule_type DNA
#residues 1-349 #label ODE
#cross-references EMBL:Z67757; NID:g1061288; PID:g1061294
REFERENCE S68433
#authors Shiozaki, K.; Russell, P.
#journal Nature (1995) 378:739-743
#title Cell-cycle control linked to extracellular environment by MAP
kinase pathway in fission yeast.
#cross-references MUID:96107317
#accession S68433
#status nucleic acid sequence not shown
#molecule_type mRNA
#residues 1-349 #label SHI
#cross-references GB:U26739; NID:g1022684; PID:g1022685
GENETICS
#gene STY1; SPC1; PHH1
#map_position 1L
CLASSIFICATION #superfamily kinase-related transforming protein; protein
kinase homology
KEYWORDS ATP; phosphoprotein; phosphotransferase; protein kinase
FEATURE
18-268 #domain protein kinase homology #label KIN\
26-34 #region protein kinase ATP-binding motif\
171 #binding_site phosphate (Thr) (covalent) #status
predicted\
173 #binding_site phosphate (Tyr) (covalent) #status
predicted\

SUMMARY #length 349 #molecular-weight 40222 #checksum 4702
Query Match 100.0%; Score 19; DB 2; Length 349;
Best Local Similarity 25.0%; Pred. No. 1.53e+02;
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 325 PVETWKVM 332
QY 2 PXXXXXXM 9

RESULT 10 S44261 #type complete
ENTRY SRG1 protein - Arabidopsis thaliana
TITLE #formal name Arabidopsis thaliana #common_name mouse-ear
ORGANISM cress
DATE 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change
12-Feb-1999
ACCESSIONS S44261
REFERENCE #authors Callard, D.; Axelos, M.; Mazzolini, L.
#submission submitted to the EMBL Data Library, April 1994
#accession S44261
#molecule_type mRNA
#residues 1-358 #label CAL
#cross-references EMBL:X79052; NID:g479046; PID:g479047
GENETICS
#gene SRG1
CLASSIFICATION #superfamily 1-aminocyclopropane-1-carboxylate oxidase
SUMMARY #length 358 #molecular-weight 41039 #checksum 7200

Query Match 100.0%; Score 19; DB 2; Length 358;
Best Local Similarity 25.0%; Pred. No. 1.53e+02;
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 166 PFRDTLEM 173
QY 2 PXXXXXXM 9

RESULT 11 I39525 #type complete
ENTRY fructose-bisphosphatase (EC 3.1.3.11) - Alcaligenes eutrophus
TITLE plasmid pHG1
ALTERNATE_NAMES fructose-1,6-bisphosphate/sedoheptulose-1,7-bisphosphate
phosphatase
ORGANISM #formal name Alcaligenes eutrophus
DATE 19-Jul-1996 #sequence_revision 19-Jul-1996 #text_change
20-Mar-1998
ACCESSIONS I39525; PQ0044
REFERENCE I39525
#authors Yoo, J.; Bowien, B.
#journal Microbiology (1995) 31:55-61
#title Analysis of the cbf genes from Alcaligenes eutrophus that
encode fructose-1,6-/sedoheptulase-1,7-bisphosphatase.
#accession I39525
#status preliminary; translated from GB/EMBL/DBJ
#molecule_type DNA
#residues 1-364 #label RES
#cross-references EMBL:U16792; NID:g901866; PID:g901867
#experimental_source strain H16
REFERENCE JQ0399
#authors Kossmann, J.; Klintworth, R.; Bowien, B.
#journal Gene (1989) 85:247-252
#title Sequence analysis of the chromosomal and plasmid genes
encoding phosphoribulokinase from Alcaligenes eutrophus.
#cross-references MUID:90152372
#accession PQ0044
#molecule_type DNA
#residues 159-364 #label KOS
#cross-references GB:M33562; NID:gl41905; PID:gl41906
#experimental_source strain H16

```

```

subtilis
#formal_name Bacillus subtilis
#sequence_revision 05-Dec-1997 #text_change
05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change
24-Sep-1998
H70046
ACCESSIONS
REFERENCE
#authors
Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.;
Alloni, G.; Azevedo, V.; Bertero, M.G.; Bessieres, P.;
Bolotin, A.; Borchert, S.; Boriss, R.; Boursier, L.; Brans,
A.; Braun, M.; Brignelli, S.C.; Bron, S.; Brouillet, S.;
Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
Choi, S.K.; Codani, J.J.; Conerton, I.F.; Cummings, N.J.;
Daniel, R.A.; Denizot, F.; Devine, K.M.; Duesterhoeft, A.;
Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.;
Fabret, C.; Ferrari, E.; Foulger, D.; Fritz, C.; Fujita,
M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galleron, N.; Ghim,
S.Y.; Glaser, P.; Goffeau, A.; Golightly, E.J.; Grandi, G.;
Guiseppi, G.; Guy, B.J.; Haga, K.; Halech, J.; Harwood,
C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.;
Hullo, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karamata, D.;
Kashara, Y.; Klaerr-Blanchard, M.; Klein, C.; Kobayashi,
Y.; Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.;
Kurita, K.; Lapidus, A.; Lardinois, S.; Lauber, J.;
Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.;
Maueel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno,
M.; Moestli, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly,
M.; Ogawa, K.; Ogihara, A.; Oudega, B.; Park, S.H.; Parro,
V.; Pohl, T.M.; Portetalle, D.; Porwollik, S.; Prescott,
A.M.; Presecan, E.; Pujic, P.; Purnelle, B.; Rappoport, G.;
Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.;
Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon, E.;
Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.;
Sekowska, A.; Seror, S.J.; Serior, P.; Shin, B.S.; Soldo,
B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Takahashi, H.;
Takemaru, K.; Takeuchi, M.; Tanakoshi, A.; Tanaka, T.;
Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.;
Vandenbol, M.; Vannier, F.; Vassarotti, A.; Viari, A.;
Wambutt, R.; Wedler, E.; Wedler, H.; Weitenecker, T.;
Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto,
K.; Yata, K.; Yoshida, K.; Yoshikawa, H.F.; Zumstein, E.;
Yoshikawa, H.; Danchin, A.
Nature (1997) 390:249-256
The complete genome sequence of the Gram-positive bacterium
Bacillus subtilis.
#cross-references MUID:98044033
#accession H70046
#status preliminary; nucleic acid sequence not shown;
translation not shown
#molecule_type DNA
#residues 1-263 #label KUN
#cross-references GB:Z99120; GB:Z99121; GB:AL009126; NID:g2635827;
PID:e1186007; PID:g2635832; NID:g2635613;
PID:e1184398; PID:g2635816
#experimental_source strain 168
GENETICS
#gene
#classification #superfamily ribitol dehydrogenase; short-chain alcohol
dehydrogenase homology
FEATURE
8-185 #domain short-chain alcohol dehydrogenase homology
#label SADH
SUMMARY #length 263 #molecular-weight 28222 #checksum 1280
Query Match 100.0%; Score 19; DB 2; Length 263;
Best Local Similarity 25.0%; Pred.No. 1.53e+02;
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
Db 149 PTMIPYSM 156
QY 2 PXXXXXXM 9
RESULT 7
#gene
#genetics
#comment This is one of the proteins, coded by nodulation genes, that are
required for this bacterium to invade and stimulate nodule
formation in its hosts. It is involved in mediating the
host-specific activation of the nodABC genes.

```

```

D64206 #type complete
ribose-phosphate pyrophosphokinase (EC 2.7.6.1) - Mycoplasma
genitalium (SGC3)
phosphoribosylpyrophosphate synthetase
#formal_name Mycoplasma genitalium
#sequence_revision 17-Nov-1995 #text_change
13-Sep-1998
ACCESSIONS
REFERENCE
#authors
Fraser, C.M.; Gocayne, J.D.; White, O.; Adams, M.D.; Clayton,
R.A.; Fleischmann, R.D.; Bult, C.J.; Kerlavage, A.R.;
Sutton, G.; Kelley, J.M.; Fritchman, J.L.; Weidman, J.F.;
Small, K.V.; Sandusky, M.; Fuhrmann, J.; Nguyen, D.;
Uterback, T.R.; Saudek, D.M.; Phillips, C.A.; Merrick,
J.M.; Tomb, J.F.; Dougherty, B.A.; Bott, K.F.; Hu, P.C.;
Lucier, T.S.; Peterson, S.N.; Smith, H.O.; Hutchison III,
C.A.; Venter, J.C.
Science (1995) 270:397-403
The minimal gene complement of Mycoplasma genitalium.
#cross-references MUID:96026346
#accession D64206
#status preliminary; nucleic acid sequence not shown;
translation not shown
#molecule_type DNA
#residues 1-297 #label TIGR
#cross-references GB:U39685; GB:L43967; NID:gl045729; PID:gl045732;
TIGR:MG058
#experimental_source strain G-37
GENETICS
#genetic_code SGC3
#function nucleotide biosynthesis
#pathway #superfamily ribose-phosphate pyrophosphokinase catalytic
chain
KEYWORDS diphosphotransferase; magnesium; nucleotide biosynthesis
FEATURE
186-198 #region phosphoribosylpyrophosphate binding\
99,101,110,114 #binding_site magnesium (Asp, His, Asp, Asp) #status
predicted
SUMMARY #length 297 #molecular-weight 33556 #checksum 2532
Query Match 100.0%; Score 19; DB 2; Length 297;
Best Local Similarity 25.0%; Pred.No. 1.53e+02;
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
Db 32 PNVDLSM 39
QY 2 PXXXXXXM 9
RESULT 8
#type complete
ENTRY nucleation protein nodd - Azorhizobium caulinodans
TITLE #formal_name Azorhizobium caulinodans
ORGANISM 10-Aug-1990 #sequence_revision 13-Jan-1995 #text_change
DATE 20-Feb-1998
ACCESSIONS
REFERENCE
#authors
Goethals, K.; Van den Eede, G.; Van Montagu, M.; Holsters, M.
J. Bacteriol. (1990) 172:2658-2666
#journal Identification and characterization of a functional nodd gene
in Azorhizobium caulinodans ORS571.
#cross-references MUID:90236930
#accession A35268
#molecule_type DNA
#residues 1-314 #label GOE
#cross-references GB:M60872; NID:gl52069; PID:gl52070
COMMENT This is one of the proteins, coded by nodulation genes, that are
required for this bacterium to invade and stimulate nodule
formation in its hosts. It is involved in mediating the
host-specific activation of the nodABC genes.
GENETICS
#gene nodd

```

```

kinase II) #status experimental\
#binding_site phosphate (Ser) (covalent) (by casein
kinase II) (partial) #status experimental
SUMMARY #length 224 #molecular-weight 25107 #checksum 8112

Query Match 100.0%; Score 19; DB 1; Length 224;
Best Local Similarity 25.0%; Pred. NO. 1.53e+02;
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 101 PFLOPEVM 108
|
QY 2 PXXXXXXM 9

RESULT 3
ENTRY
TITLE S19734 #type complete
ALTERNATE_NAMES glutathione transferase (EC 2.5.1.18) - chicken
ORGANISM #formal_name Gallus gallus #common_name chicken
DATE 22-Nov-1993 #sequence_revision 10-Nov-1995 #text_change
08-Sep-1997
ACCESSIONS S19734
REFERENCE S19734
#authors Chang, L.H.; Fan, J.Y.; Liu, L.F.; Tsai, S.P.; Tam, M.F.;
#journal Biochem. J. (1992) 281:545-551
#title Cloning and expression of a chick liver glutathione
S-transferase CL 3 subunit with the use of a baculovirus
expression system.
#cross-references MUID:92143826
#accession S19734
#molecule_type mRNA
#residues 1-229 #label CHA
#cross-references EMBL:M38219; NID:g211529; PID:g211530
CLASSIFICATION #superfamily glutathione transferase
KEYWORDS
SUMMARY #length 229 #molecular-weight 26326 #checksum 4524

Query Match 100.0%; Score 19; DB 2; Length 229;
Best Local Similarity 25.0%; Pred. NO. 1.53e+02;
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 56 PMVEIDGM 63
|
QY 2 PXXXXXXM 9

RESULT 4
ENTRY
TITLE S33175 #type fragment
ORGANISM hydroxymethylglutaryl-CoA reductase (NADPH) (EC 1.1.1.34) -
rat (fragment)
#formal_name Rattus norvegicus #common_name Norway rat
DATE 22-Nov-1993 #sequence_revision 10-Nov-1995 #text_change
08-Sep-1997
ACCESSIONS S33175; S11079
REFERENCE S33175
#authors Khan, S.; Kabat, S.; Stambrook, P.
#submission Submitted to the EMBL Data Library, October 1990
#description Rat HMG-CoA reductase cDNA sequence contains an unusual 36 bp
insert bounded by inverted repeats.
#accession S33175
#molecule_type mRNA
#residues 1-244 #label KHA
#cross-references EMBL:X55286; NID:g296924; PID:g296925
REFERENCE S11079
#authors Clarke, P.R.; Hardie, D.G.
#journal FEBS Lett. (1990) 269:213-217
#title Calmodulin-dependent multiprotein kinase and protein kinase C
phosphorylate the same site on HMG-CoA reductase as the
AMP-activated protein kinase.
#cross-references MUID:90353576
#accession S11079
#molecule_type protein
#residues 223-231 #label FEB
CLASSIFICATION #superfamily hydroxymethylglutaryl-CoA reductase (NADPH)

```

```

KEYWORDS NADP; oxidoreductase
SUMMARY #length 244 #checksum 2867

Query Match 100.0%; Score 19; DB 2; Length 244;
Best Local Similarity 25.0%; Pred. NO. 1.53e+02;
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 169 PQOACLOM 176
|
QY 2 PXXXXXXM 9

RESULT 5
ENTRY
TITLE LNHUP6 #type complete
ALTERNATE_NAMES pulmonary surfactant protein A precursor (clone 6A) - human
ORGANISM pulmonary surfactant 32k apoprotein; pulmonary
surfactant-associated protein PSP-A
#formal_name Homo sapiens #common_name man
DATE 31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change
18-Sep-1998
ACCESSIONS A25720
REFERENCE A25720
#authors Floros, J.; Steinbrink, R.; Jacobs, K.; Phelps, D.; Kriz, R.;
Rechy, M.; Sultzman, L.; Jones, S.; Taeusch, H.W.; Frank,
H.A.; Fritsch, E.F.
#journal J. Biol. Chem. (1986) 261:9029-9033
#title Isolation and characterization of cDNA clones for the 35-kDa
pulmonary surfactant-associated protein.
#cross-references MUID:86250832
#accession A25720
#molecule_type mRNA
#residues 1-248 #label FLO
#cross-references GB:M13686; NID:g190669; PID:g190670
#note Part of the sequence was confirmed by protein sequencing
the amino end of the mature protein, which was not
identified, is partially acetylated
#note clones corresponding to two different proteins were
sequenced. Cotranslational modifications of the
proteins (including acetylation) produce multiple
isoforms

```

```

GENETICS
#gene GDB:SFTPAL; SFTPL; SP-A; SP-AI
#cross-references GDB:I19593; OMIM:178630
#map_position 10q22-10q23
CLASSIFICATION #superfamily mannose-binding lectin; C-type lectin homology
KEYWORDS acetylated amino end; alveolar proteinosis; calcium; gaseous
exchange; glycoprotein; hydroxyproline; lung; pulmonary
surfactant; respiratory distress syndrome
FEATURE
1-20 #domain signal sequence #status predicted #label SIG\
21-248 #product pulmonary surfactant protein A #status
predicted #label MAT\
127-246 #domain C-type lectin homology #label CLEC\
21 #modified_site acetylated amino end (Glu) (in mature
form) #status predicted\
30-33,36,42,54,57, #modified_site 4-hydroxyproline (Pro) #status predicted\
63,76,79,82,91,97 #binding_site carbohydrate (Asn) (covalent) #status
predicted
SUMMARY #length 248 #molecular-weight 26214 #checksum 3326

Query Match 100.0%; Score 19; DB 1; Length 248;
Best Local Similarity 25.0%; Pred. NO. 1.53e+02;
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 59 PMGPPGEM 66
|
QY 2 PXXXXXXM 9

RESULT 6
ENTRY
TITLE H70046 #type complete
#accession H70046
#molecule_type protein
#residues 1-231
CLASSIFICATION #superfamily protein reductase homolog yvrd - Bacillus

```

Best Local Similarity 25.0%; Pred. No. 1.53e+02;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 120 PHPLSPM 127

QY 2 PXXXXXX 9

RESULT 2

ENTRY KB0A2 #type complete  
TITLE beta-casein precursor - bovine  
ORGANISM #formal\_name Bos primigenius taurus #common\_name cattle  
DATE 24-Apr-1984 #sequence\_revision 12-May-1995 #text\_change 05-Sep-1997  
ACCESSIONS I45873; B29087; S01860; A25846; S02429; A90489; A91191; B91192; C91192; D91192; A90739; A91413; A03110  
REFERENCE I45873  
#authors Bousling, J.; Ring, J.M.; Stewart, A.F.; Mackinlay, A.G.  
#journal Aust. J. Biol. Sci. (1988) 41:527-537  
#title Complete nucleotide sequence of the bovine beta-casein gene.  
#cross-references MUID:90147279  
#accession I45873  
#status preliminary; translated from GB/EMBL/DBJ  
#molecule\_type DNA  
#residues 1-81, 'H', 83-224 #label BON  
#cross-references GB:M55156; NID:g162804; PID:g162805  
REFERENCE A93062  
#authors Stewart, A.F.; Bousling, J.; Beattie, C.W.; Shah, F.; Willis, I.M.; Mackinlay, A.G.  
#journal Mol. Biol. Evol. (1987) 4:231-241  
#title Complete nucleotide sequences of bovine alpha-s2- and beta-casein cDNAs: comparisons with related sequences in other species.  
#cross-references MUID:8818898  
#accession B29087  
#status translation not shown  
#molecule\_type mRNA  
#residues 1-224 #label STE  
#cross-references GB:M16645; NID:g162930; PID:g162931  
#experimental\_source A2 variant  
REFERENCE S01860  
#authors Baev, A.A.; Smirnov, I.K.; Gorodetski, S.I.  
#journal Mol. Biol. (1987) 21:214-222  
#title Primary structure of bovine beta-casein cDNA.  
#accession S01860  
#molecule\_type mRNA  
#residues 1-81, 'H', 83-224 #label BAE  
#cross-references EMBL:X06359; NID:g171; PID:g757752  
#experimental\_source A1 variant  
#note this paper is a translation of the Russian paper published in Mol. Biol. Moscow (1987) 21: 255-265  
REFERENCE A25846  
#authors Jimenez-Flores, R.; Kang, Y.C.; Richardson, T.  
#journal Biochem. Biophys. Res. Commun. (1987) 142:617-621  
#title Cloning and sequence analysis of bovine beta-casein cDNA.  
#cross-references MUID:87128158  
#accession A25846  
#molecule\_type mRNA  
#residues 1-107, 'L', 109-151, 'PL', 154-209, 'Q', 211-224 #label JIM  
#cross-references GB:M15132; NID:g162796; PID:g162797  
REFERENCE S02429  
#authors Carles, C.; Huet, J.C.; Ribadeau-Dumas, B.  
#journal FEBS Lett. (1988) 229:265-272  
#title A new strategy for primary structure determination of proteins: application to bovine beta-casein.  
#cross-references MUID:86152252  
#accession S02429  
#molecule\_type protein  
#residues 16-81, 'H', 83-224 #label CAR  
#experimental\_source A1 variant  
REFERENCE A90489  
#authors Yan, S.B.; Wold, F.  
#journal Biochemistry (1984) 23:3759-3765

#title Neoglycoproteins: in vitro introduction of glycosyl units at glutamylase in beta-casein using transglutaminase.  
#cross-references MUID:85000478  
#accession A90489  
#molecule\_type protein  
#residues 16-224 #label YAN  
REFERENCE A91191  
#authors Ribadeau-Dumas, B.; Brignon, G.; Grosclaude, F.; Mercier, J.C.  
#journal Eur. J. Biochem. (1972) 25:505-514  
#title Structure primaire de la caseine bovine.  
#cross-references MUID:7233212  
#accession A91191  
#molecule\_type protein  
#residues 16-131, 'Q', 133-151, 'PL', 154-189, 'E', 191-209, 'Q', 211-224 #label RIB  
#experimental\_source A2 variant  
#note article in French with an English abstract  
REFERENCE A91192  
#authors Grosclaude, F.; Mahe, M.F.; Mercier, J.C.; Ribadeau-Dumas, B.  
#journal Eur. J. Biochem. (1972) 26:328-337  
#title Caracterisation des variants genetiques des caseines alpha-S1 et beta bovines.  
#cross-references MUID:7214259  
#note article in French with an English abstract  
#accession B91192  
#molecule\_type protein  
#residues 16-81, 'H', 83-131, 'Q', 133-151, 'PL', 154-189, 'E', 191-209, 'Q', 211-224 #label VAI  
#experimental\_source A1 variant  
#accession C91192  
#molecule\_type protein  
#residues 16-81, 'H', 83-131, 'Q', 133-136, 'R', 138-151, 'PL', 154-189, 'E', 191-209, 'Q', 211-224 #label VAB  
#experimental\_source B variant  
#accession D91192  
#molecule\_type protein  
#residues 16-51, 'K', 53-81, 'H', 83-131, 'Q', 133-151, 'PL', 154-189, 'E', 191-209, 'Q', 211-224 #label VAC  
#experimental\_source C variant  
#note this variant lacks a phosphate group on 50-Ser  
REFERENCE A90739  
#authors Ribadeau-Dumas, B.; Grosclaude, F.; Mercier, J.C.  
#journal C. R. Acad. Sci. D Sci. Nat. (1970) 270:2369-2372  
#title Localisation dans la chaine peptidique de la caseine beta bovine de la substitution His/Gln differenciant les variants genetiques A2 et A3.  
#note article in French with an English abstract  
#accession A90739  
#molecule\_type protein  
#residues 118-120, 'Q', 122-124 #label VA3  
#experimental\_source A3 variant  
REFERENCE A91413  
#authors Grosclaude, F.; Mahe, M.F.; Voglino, G.F.  
#journal FEBS Lett. (1974) 45:3-5  
#title Le variant beta-E et le code de phosphorylation des caseines bovines.  
#cross-references MUID:75005247  
#note article in French with an English abstract  
#accession A91413  
#molecule\_type protein  
#residues 48-50, 'K', 52-63 #label VAE  
#experimental\_source E variant  
#note 50-Ser is phosphorylated  
#note The sequence shown is the A2 variant.  
COMMENT  
GENETICS  
#introns 17/3; 26/3; 35/3; 43/3; 57/3; 223/3  
CLASSIFICATION #superfamily beta-casein  
KEYWORDS milk; phosphoprotein  
FEATURE  
1-15  
16-224  
30,32,33,34  
#domain signal sequence #status predicted #label SIG\  
#product beta-casein #status experimental #label MAT\  
#binding\_site phosphate (Ser) (covalent) (by casein

\*\*\*\*\*



(TM)

\*\*\*\*\*

Release 3.1A John F. Collins, BioComputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Sat Apr 15 02:17:47 2000; MasPar time 3.19 Seconds  
113.132 Million cell updates/sec  
Tabular output not generated.

Title: >US-08-452-843-30  
Description: (1-9) from US08452843.pep  
Perfect Score: 19  
Sequence: 1 XPXXXXXXM 9

Scoring table: PAM 150  
Gap 15

Searched: 122810 seqs, 40068593 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: pir62  
1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 11.700; Variance 4.525; scale 2.586

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	19	100.0	192	1 KXSHA	kappa-casein precursor	1.53e+02
2	19	100.0	224	1 KBROA2	beta-casein precursor	1.53e+02
3	19	100.0	229	2 S19734	glutathione transferase	1.53e+02
4	19	100.0	244	2 S33175	hydroxymethylglutaryl	1.53e+02
5	19	100.0	248	1 LNHUP6	pulmonary surfactant	1.53e+02
6	19	100.0	263	2 H70046	ketoacyl-carrier prot	1.53e+02
7	19	100.0	297	2 D64206	ribose-phosphate pyro	1.53e+02
8	19	100.0	314	1 A35268	nucleation protein no	1.53e+02
9	19	100.0	349	2 S68675	mitogen-activated pro	1.53e+02
10	19	100.0	358	2 S44261	SRG1 protein - Arabid	1.53e+02
11	19	100.0	364	2 I39525	fructose-bisphosphata	1.53e+02
12	19	100.0	368	2 S57273	lignin peroxidase (EC	1.53e+02
13	19	100.0	370	2 B43673	chloromuconate cyclo	1.53e+02
14	19	100.0	394	1 KIBSGM	phosphoglycerate kina	1.53e+02
15	19	100.0	395	1 A46345	gene III protein - ph	1.53e+02
16	19	100.0	401	1 TWTGTG	phosphoglycerate kina	1.53e+02
17	19	100.0	420	2 A57742	cyclin Ia - maize	1.53e+02
18	19	100.0	429	2 A36220	transforming protein o	1.53e+02
19	19	100.0	433	1 E64242	GTP-binding protein o	1.53e+02
20	19	100.0	445	1 R58SMT	probable site-specifi	1.53e+02
21	19	100.0	468	2 A37176	glutamate--ammonia li	1.53e+02
22	19	100.0	510	2 S15893	triacylglycerol lipas	1.53e+02
23	19	100.0	529	2 F70550	probable accD1 protei	1.53e+02

24	19	100.0	562	2 S27800	elastase precursor -	1.53e+02
25	19	100.0	580	1 ERECEX	gamma-glutamyltransfe	1.53e+02
26	19	100.0	593	1 XYZFG	fzG protein - Myxoco	1.53e+02
27	19	100.0	654	2 B55729	biotin carboxyl carri	1.53e+02
28	19	100.0	662	1 A31349	arachidonate 15-lipo	1.53e+02
29	19	100.0	663	1 A38283	neurotrophic receptor	1.53e+02
30	19	100.0	685	2 A48289	thimet oligopeptidase	1.53e+02
31	19	100.0	689	1 HYHUTH	copper-transporting A	1.53e+02
32	19	100.0	708	1 JC2465	probable copper-trans	1.53e+02
33	19	100.0	731	1 JC2464	Kell blood group prot	1.53e+02
34	19	100.0	732	1 RGHUK	regulatory protein SW	1.53e+02
35	19	100.0	803	1 RBYW6	S-receptor kinase (EC	1.53e+02
36	19	100.0	824	2 S50767	alanine--tRNA ligase	1.53e+02
37	19	100.0	892	2 D64370	protein-tyrosine-phos	1.53e+02
38	19	100.0	1118	1 A49724	internal vision prote	1.53e+02
39	19	100.0	1318	1 HIBPD7	hepatocyte growth fac	1.53e+02
40	19	100.0	1375	2 JC5148	DNA-directed DNA poly	1.53e+02
41	19	100.0	1505	2 S28079	complement C3 precurs	1.53e+02
42	19	100.0	1652	2 I50711	epidermal growth fact	1.53e+02
43	19	100.0	1717	2 A45558	leukocyte antigen-rel	1.53e+02
44	19	100.0	1897	1 TDHULK	probable GTPase-activ	1.53e+02
45	19	100.0	3079	1 RGYI2		

ALIGNMENTS

RESULT	1	KKSHA	#type complete
ENTRY		kappa-casein precursor - sheep	
TITLE		#formal_name Ovis orientalis aries, Ovis ammon aries	
ORGANISM		#common_name domestic sheep	
DATE		24-Apr-1984 #sequence_revision 30-Sep-1991 #text_change 05-Sep-1997	
ACCESSIONS		S14711; A03113; A90597; S08655	
REFERENCE		S14711	
#authors		Furet, J.P.; Mercier, J.C.; Soulier, S.; Gaye, P.; Hue-Delahaie, D.; Vilotte, J.L.	
#journal		Nucleic Acids Res. (1990) 18:5286	
#title		Nucleotide sequence of ovine kappa-casein cDNA.	
#cross-references		EMBL:X51822; NID:g1293; PID:g1294	
#accession		S14711	
#molecule_type		mRNA	
#residues		1-192	#label FUR
#cross-references		EMBL:X51822; NID:g1293; PID:g1294	
REFERENCE		A91221	
#authors		Jolles, J.; Schoentgen, F.; Hermann, J.; Alais, C.; Jolles, P.	
#journal		Eur. J. Biochem. (1974) 46:127-132	
#title		The sequence of sheep kappa-casein: primary structure of para-kappa-A-casein.	
#cross-references		MUID:74309256	
#accession		A03113	
#molecule_type		protein	
#residues		22,'Q','24-27','E','29-129	#label JOL1
REFERENCE		A90597	
#authors		Jolles, J.; Fiat, A.M.; Schoentgen, F.; Alais, C.; Jolles, P.	
#journal		Biochim. Biophys. Acta (1974) 365:335-343	
#title		The amino acid sequence of sheep kappa-A-casein. II. Sequence studies concerning the kappa-A-caseinoglycopeptide and establishment of the complete primary structure of the protein.	
#cross-references		MUID:75036120	
#accession		A90597	
#molecule_type		protein	
#residues		127-192	#label JOL2
CLASSIFICATION		#superfamily kappa-casein	
KEYWORDS		glycoprotein; mammary gland; milk; phosphoprotein	
FEATURE		1-21	#domain signal sequence #status predicted #label SIGV
22-192			#product kappa-casein #status experimental #label MAT
SUMMARY		#length 192 #molecular-weight 21438 #checksum 9739	
Query Match		100.0%; Score 19; DB 1; Length 192;	



Job time : 90 secs.

O28627;  
AC 01-JAN-1998 (TRENBLrel. 05, Created)  
DT 01-JAN-1998 (TRENBLrel. 05, Last sequence update)  
DT 01-AUG-1998 (TRENBLrel. 07, Last annotation update)  
DE CONSERVED HYPOTHETICAL PROTEIN.  
GN AF1646.  
OS Archaeoglobus fulgidus.  
OC Archaea; Euryarchaeota; Archaeoglobales; Archaeoglobaceae;  
OC Archaeoglobus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-VC-16 / DSM 4304 / ATCC 49558;  
RX MEDLINE; 98049343.  
RA KLENK H.-P., CLAYTON R.A., TOMB J.-F., WHITE O., NELSON K.E.,  
RA KETCHUM K.A., DODSON R.J., GWINN M., HICKEY E.K., PETERSON J.D.,  
RA RICHARDSON D.L., KERLAVAGE A.R., GRAHAM D.E., KYRIDES N.C.,  
RA FLEISCHMANN R.D., QUACKENBUSH J., LEE N.H., SUTTON G.G., GILL S.,  
RA KIRKNESS E.F., DOUGHERTY B.A., MCKENNEY K., ADAMS M.D., LOFTUS B.,  
RA PETERSON S., REICH C.I., MCNEIL L.K., BADGER J.H., GLODEK A., ZHOU L.,  
RA OVERBEK R., GOCAYNE J.D., WEIDMAN J.F., McDONALD L., UTTERBACK T.,  
RA COTTON M.D., SPRIGGS T., ARTIACH P., KAINE B.P., SYKES S.M.,  
RA SADOW P.W., D'ANDREA K.P., BOWMAN C., FUJII C., GARLAND S.A.,  
RA MASON T.M., OLSEN G.J., FRASER C.M., SMITH H.O., WOESE C.R.,  
RA VENTER J.C.;  
RT "The complete genome sequence of the hyperthermophilic, sulphate-  
RT reducing archaeon Archaeoglobus fulgidus.";  
RL Nature 390:364-370(1997).  
DR EMBL; AE000989; AAB89596.1; -.  
DR TIGR; AF1646; -.  
KW Hypothetical protein.  
SQ SEQUENCE 261 AA; 27976 MW; 2C5D3877 CRC32;  
  
Query Match 100.0%; Score 19; DB 1; Length 261;  
Best Local Similarity 25.0%; Pred. No. 5.59e+00;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
  
Db 75 PFGCGFGM 82  
|  
Qy 2 PXXXXXXM 9  
  
RESULT 15  
ID P96189 PRELIMINARY; PRT; 274 AA.  
AC P96189;  
DT 01-MAY-1997 (TRENBLrel. 03, Created)  
DT 01-MAY-1997 (TRENBLrel. 03, Last sequence update)  
DT 01-NOV-1998 (TRENBLrel. 08, Last annotation update)  
DE HYPOTHETICAL 30.2 KD PROTEIN.  
OS Xanthomonas campestris.  
OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;  
OC Xanthomonas.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-ATCC 11645;  
RX MEDLINE; 97276896.  
RA GOMEZ P., RIBAS-APARICIO R.M., PELAEZ A.I., GOMEZ A., RODICIO M.R.;  
RT "Isolation and nucleotide sequence of the gene encoding the xanI DNA  
RT methyltransferase of Xanthomonas campestris pv. amaranthicola.";  
RL Biochim. Biophys. Acta 1351:261-266(1997).  
DR EMBL; U77781; AAD13687.1; -.  
KW Hypothetical protein.  
SQ SEQUENCE 274 AA; 30170 MW; 57A01D29 CRC32;  
  
Query Match 100.0%; Score 19; DB 2; Length 274;  
Best Local Similarity 25.0%; Pred. No. 5.59e+00;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
  
Db 129 PWMETGGM 136  
|  
Qy 2 PXXXXXXM 9

Search completed: Sat Apr 15 02:20:54 2000



```

RA NELSON K., QUACKENBUSH J., ZHOU L., KIRKNESS E.F., PETERSON S.,
RA LOFTUS B., RICHARDSON D., DOSON R., KHALAK H.G., GLODEK A.,
RA MCKENNEY K., FITZGERALD L.M., LEE N., ADAMS M.D., HICKEY E.K.,
RA BERG D.E., GOCAYNE J.D., UTTERBACK T.R., PETERSON J.D., KELLEY J.M.,
RA COTTON M.D., WEIDMAN J.M., FUJII C., BOWMAN C., WATTHEY L., WALLIN E.,
RA HAYES W.S., BORODOVSKY M., KARP P.D., SMITH H.O., FRASER C.M.,
RA VENTER J.C.;
RT "The complete genome sequence of the gastric pathogen Helicobacter
RT pylori.";
RL Nature 388:539-547(1997).
DR EMBL: AE000545; AAD07331.1; -.
DR TIGR: HP0265; -.
KW Hypothetical protein.
SQ SEQUENCE 240 AA; 26522 MW; 4FA35C4C CRC32;

Query Match 100.0%; Score 19; DB 2; Length 240;
Best Local Similarity 25.0%; Pred. No. 5.59e+00;
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 186 PFLVALM 193
|
Qy 2 PXXXXXXM 9

RESULT 11
ID P73217 PRELIMINARY; PRT; 243 AA.
AC P73217;
DT 01-FEB-1997 (TREMBlrel. 02, Created)
DT 01-FEB-1997 (TREMBlrel. 02, Last sequence update)
DT 01-JAN-1999 (TREMBlrel. 09, Last annotation update)
DE HYPOTHETICAL 26.6 KD PROTEIN.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-PCC6803;
RA TABATA S.;
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
[2]
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-PCC6803;
RX MEDLINE; 97061201.
RA KANEKO T., SATO S., KOTANI H., TANAKA A., ASAMIZU E., NAKAMURA Y.,
RA MIYAJIMA N., HIROSAWA M., SUGIURA M., SASAMOTO S., KIMURA T.,
RA HOSOUCHI T., MATSUNO A., MURAKI A., NAKAZAKI N., NARUO K., OKUMURA S.,
RA SHIMPO S., TAKEUCHI C., WADA T., WATANABE A., YAMADA M., YASUDA M.,
RA TABATA S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions.";
RL DNA Res. 3:109-136(1996).
DR EMBL: D90904; BAA17244.1; -.
KW Hypothetical protein.
SQ SEQUENCE 243 AA; 26620 MW; 784CF595 CRC32;

Query Match 100.0%; Score 19; DB 2; Length 243;
Best Local Similarity 25.0%; Pred. No. 5.59e+00;
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 221 PQAVQNM 228
|
Qy 2 PXXXXXXM 9

RESULT 12
ID P72952 PRELIMINARY; PRT; 255 AA.
AC P72952;
DT 01-FEB-1997 (TREMBlrel. 02, Created)
DT 01-FEB-1997 (TREMBlrel. 02, Last sequence update)
DT 01-JAN-1999 (TREMBlrel. 09, Last annotation update)
DE HYPOTHETICAL 27.1 KD PROTEIN.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.

```

```

RN SEQUENCE FROM N.A.
RC STRAIN-PCC6803;
RA TABATA S.;
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
[2]
RN SEQUENCE FROM N.A.
RC STRAIN-PCC6803;
RX MEDLINE; 97061201.
RA KANEKO T., SATO S., KOTANI H., TANAKA A., ASAMIZU E., NAKAMURA Y.,
RA MIYAJIMA N., HIROSAWA M., SUGIURA M., SASAMOTO S., KIMURA T.,
RA HOSOUCHI T., MATSUNO A., MURAKI A., NAKAZAKI N., NARUO K., OKUMURA S.,
RA SHIMPO S., TAKEUCHI C., WADA T., WATANABE A., YAMADA M., YASUDA M.,
RA TABATA S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions.";
RL DNA Res. 3:109-136(1996).
DR EMBL: D90902; BAA16970.1; -.
KW Hypothetical protein.
SQ SEQUENCE 255 AA; 27096 MW; 629A941D CRC32;

Query Match 100.0%; Score 19; DB 2; Length 255;
Best Local Similarity 25.0%; Pred. No. 5.59e+00;
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 240 PVVECCQM 247
|
Qy 2 PXXXXXXM 9

RESULT 13
ID P72950 PRELIMINARY; PRT; 256 AA.
AC P72950;
DT 01-FEB-1997 (TREMBlrel. 02, Created)
DT 01-FEB-1997 (TREMBlrel. 02, Last sequence update)
DT 01-JAN-1999 (TREMBlrel. 09, Last annotation update)
DE HYPOTHETICAL 28.7 KD PROTEIN.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-PCC6803;
RA TABATA S.;
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
[2]
RN SEQUENCE FROM N.A.
RC STRAIN-PCC6803;
RX MEDLINE; 97061201.
RA KANEKO T., SATO S., KOTANI H., TANAKA A., ASAMIZU E., NAKAMURA Y.,
RA MIYAJIMA N., HIROSAWA M., SUGIURA M., SASAMOTO S., KIMURA T.,
RA HOSOUCHI T., MATSUNO A., MURAKI A., NAKAZAKI N., NARUO K., OKUMURA S.,
RA SHIMPO S., TAKEUCHI C., WADA T., WATANABE A., YAMADA M., YASUDA M.,
RA TABATA S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions.";
RL DNA Res. 3:109-136(1996).
DR EMBL: D90902; BAA16968.1; -.
KW Hypothetical protein.
SQ SEQUENCE 256 AA; 28730 MW; 72D38E74 CRC32;

Query Match 100.0%; Score 19; DB 2; Length 256;
Best Local Similarity 25.0%; Pred. No. 5.59e+00;
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 218 PAAIKAKM 225
|
Qy 2 PXXXXXXM 9

RESULT 14
ID O28627 PRELIMINARY; PRT; 261 AA.

```

DT 01-JAN-1999 (TREMBlrel. 09, Last annotation update)  
DE 217AA LONG HYPOTHETICAL AROM PROTEIN.  
GN PH1049.  
OS Pyrococcus horikoshii.  
OC Archaea: Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=OT3;  
RX MEDLINE: 98344137.  
RA KAWARABAYASI Y., SAWADA M., HORIKAWA H., HAIKAWA Y., HINO Y.,  
RA YAMAMOTO S., SEKINE M., BABA S., KOSUGI H., HOSOIYAMA A., NAGAI Y.,  
RA SAKAI M., OGURA K., OTUKA R., NAKAZAWA H., TAKAMIYA M., OHFUKU Y.,  
RA FUNAHASHI T., TANAKA T., KUDOH Y., YAMAZAKI J., KUSHIDA N., OGUCHI A.,  
RA AOKI K., NAKAMURA Y., ROBB T.F., HORIKOSHI K., MASUCHI Y., SHIZUYA H.,  
RA KIKUCHI H.;  
RT "Complete sequence and gene organization of the genome of a hyper-  
thermophilic archaeobacterium, Pyrococcus horikoshii OT3.";  
RL DNA Res. 5:55-76(1998).  
DR EMBL: AF000004; BAA30147.1; -. 70619188 CRC32;  
SQ SEQUENCE 217 AA; 24303 MW; 70619188 CRC32;  
  
Query Match 100.0%; Score 19; DB 1; Length 217;  
Best Local Similarity 25.0%; Pred. No. 5.59e+00;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
  
Db 131 PDPKQKEM 138  
QY 2 PXXXXXXM 9  
  
RESULT 7  
ID P95933 PRELIMINARY; PRT; 222 AA.  
AC P95933;  
DT 01-MAY-1997 (TREMBlrel. 03, Created)  
DT 01-MAY-1997 (TREMBlrel. 03, Last sequence update)  
DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)  
DE ORF C01035.  
OS Sulfolobus solfataricus.  
OC Archaea: Crenarchaeota; Sulfolobales; Sulfolobus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=P2;  
RX MEDLINE: 97055432.  
RA SENEN C.W., KLENG F., SINGH R.K., ALLARD G., CHAN C.C.Y., LIU Q.Y.,  
RA PENNY S.L., YOUNG F., SCHENK M.E., GAASTERLAND T., DOOLITTLE W.F.,  
RA RAGAN M.A., CHARLEBOIS R.L.;  
RT "Organizational characteristics and information content of an archaeal  
genome: 156 kb of sequence from Sulfolobus solfataricus P2.";  
RL Mol. Microbiol. 22:175-191(1996).  
DR EMBL: Y08256; CAA69437.1; -.  
DR PFAM: PF00702; Hydrolase: 1.  
SQ SEQUENCE 222 AA; 25448 MW; 3BC8793F CRC32;  
  
Query Match 100.0%; Score 19; DB 1; Length 222;  
Best Local Similarity 25.0%; Pred. No. 5.59e+00;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
  
Db 153 PKIFSYAM 160  
QY 2 PXXXXXXM 9  
  
RESULT 8  
ID P95962 PRELIMINARY; PRT; 227 AA.  
AC P95962;  
DT 01-MAY-1997 (TREMBlrel. 03, Created)  
DT 01-MAY-1997 (TREMBlrel. 03, Last sequence update)  
DT 01-JUN-1998 (TREMBlrel. 06, Last annotation update)  
DE ORF C04023.  
OS Sulfolobus solfataricus.  
OC Archaea: Crenarchaeota; Sulfolobales; Sulfolobus.  
RN [1]  
RP SEQUENCE FROM N.A.

RC STRAIN=P2;  
RX MEDLINE: 97055432.  
RA SENEN C.W., KLENG F., SINGH R.K., ALLARD G., CHAN C.C.Y., LIU Q.Y.,  
RA PENNY S.L., YOUNG F., SCHENK M.E., GAASTERLAND T., DOOLITTLE W.F.,  
RA RAGAN M.A., CHARLEBOIS R.L.;  
RT "Organizational characteristics and information content of an archaeal  
genome: 156 kb of sequence from Sulfolobus solfataricus P2.";  
RL Mol. Microbiol. 22:175-191(1996).  
DR EMBL: Y08257; CAA69552.1; -.  
SQ SEQUENCE 227 AA; 25108 MW; 3438773D CRC32;  
  
Query Match 100.0%; Score 19; DB 1; Length 227;  
Best Local Similarity 25.0%; Pred. No. 5.59e+00;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
  
Db 162 PIGGRFTM 169  
QY 2 PXXXXXXM 9  
  
RESULT 9  
ID Q9YEN3 PRELIMINARY; PRT; 235 AA.  
AC Q9YEN3;  
DT 01-NOV-1999 (TREMBlrel. 12, Created)  
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)  
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)  
DE 235AA LONG HYPOTHETICAL URACIL PHOSPHORIBOSYLTRANSFERASE.  
GN APE0545.  
OS Aeropyrum pernix.  
OC Archaea: Crenarchaeota; Aeropyrum.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K1.  
RX MEDLINE: 99310339.  
RA KAWARABAYASI Y., HINO Y., HORIKAWA H., YAMAZAKI S., HAIKAWA Y.,  
RA JIN-NO K., TAKAHASHI M., SEKINE M., BABA S., ANKAI A., KOSUGI H.,  
RA HOSOIYAMA A., FUKUI S., NAGAI Y., NISHIJIMA K., NAKAZAWA H.,  
RA TAKAMIYA M., MASUDA S., FUNAHASHI T., TANAKA T., KUDOH Y.,  
RA YAMAZAKI J., KUSHIDA N., OGUCHI A., AOKI K., KUBOTA K., NAKAMURA Y.,  
RA NOMURA N., SAKO Y., KIKUCHI H.;  
RT "Complete genome sequence of an aerobic hyper-thermophilic  
Crenarchaeon, Aeropyrum pernix K1.";  
RL DNA Res. 6:83-101(1999).  
DR EMBL: AP000060; BAA79513.1; -.  
KW Transferase: Glycosyltransferase.  
SQ SEQUENCE 235 AA; 25513 MW; 15E46B52 CRC32;  
  
Query Match 100.0%; Score 19; DB 1; Length 235;  
Best Local Similarity 25.0%; Pred. No. 5.59e+00;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
  
Db 151 PAVLDPM 158  
QY 2 PXXXXXXM 9  
  
RESULT 10  
ID O25044 PRELIMINARY; PRT; 240 AA.  
AC O25044;  
DT 01-JAN-1998 (TREMBlrel. 05, Created)  
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)  
DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)  
DE CYTOCHROME C BIOGENESIS PROTEIN (CCDA).  
GN HP0265.  
OS Helicobacter pylori (Campylobacter pylori).  
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;  
OC Helicobacter.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=26695;  
RX MEDLINE: 97394467.  
RA TOMB J.-F., WHITE O., KERLAVAGE A.R., CLAYTON R.A., SUTTON G.G.,  
RA FLEISCHMANN R.D., KETCHUM K.A., KLENK H.-P., GILL S., DOUGHERTY B.A.,

RC STRAIN-K-12;  
RA BLATTNER F.R., PLUNKETT G. III, MAYHEW G.F., PERNA N.T., GLASNER F.D.;  
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.

RA ROBERTS D., ALLEN E., ARAUJO R., APARICIO A., CHUNG E., DAVIS K.,  
RA DUNCAN M., FEDERSPIEL N., HYMAN R., KALMAN S., KOMP C., KURDI O.,  
RA LEW H., LIN D., NAMATH A., OFENER P., SCHRAMM S., DAVIS R.W.;  
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AE000156; AAC73504.1; -;  
DR EMBL; U82664; BAB40255.1; -;  
SQ SEQUENCE 135 AA; 15304 MW; D312A7DA CRC32;

Query Match 100.0%; Score 19; DB 2; Length 135;  
Best Local Similarity 25.0%; Pred. No. 5.59e+00;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 6 PKQEVTLM 13  
|  
QY 2 PXXXXXXM 9

RESULT 3 PRELIMINARY; PRT; 202 AA.  
ID P73219  
AC P73219;  
DT 01-FEB-1997 (TREMBlrel. 02, Created)  
DT 01-FEB-1997 (TREMBlrel. 02, Last sequence update)  
DT 01-JAN-1999 (TREMBlrel. 09, Last annotation update)  
DE FIBRILLIN

OS Synechocystis sp. (strain PCC 6803);  
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.  
RN [1]

RP SEQUENCE FROM N.A.  
RC STRAIN-PCC6803;  
RA TABATA S.;  
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.  
RN [2]

RP SEQUENCE FROM N.A.  
RC STRAIN-PCC6803;  
RX MEDLINE; 97061201.  
RA KANEKO T., SATO S., KOTANI H., TANAKA A., ASAMIZU E., NAKAMURA Y.,  
RA MIYAJIMA N., HIROSAWA M., SUGIURA M., SASAMOTO S., KIMURA T.,  
RA HOSOUCHI T., MATSUNO A., MURAKI A., NAKAZAKI N., NARUO K., OKUMURA S.,  
RA SHIMO S., TAKEUCHI C., WADA T., WATANABE A., YAMADA M., YASUDA M.,  
RA TABATA S.;  
RL "Sequence analysis of the genome of the unicellular cyanobacterium  
RT Synechocystis sp. strain PCC6803. II. Sequence determination of the  
RT entire genome and assignment of potential protein-coding regions."  
RL DNA Res. 3:109-136(1996).  
DR EMBL; D90904; BAA17246.1; -;  
SQ SEQUENCE 202 AA; 22728 MW; E68453F6 CRC32;

Query Match 100.0%; Score 19; DB 2; Length 202;  
Best Local Similarity 25.0%; Pred. No. 5.59e+00;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 2 PMSMDANN 9  
|  
QY 2 PXXXXXXM 9

RESULT 4 PRELIMINARY; PRT; 211 AA.  
ID P74351  
AC P74351;  
DT 01-FEB-1997 (TREMBlrel. 02, Created)  
DT 01-FEB-1997 (TREMBlrel. 02, Last sequence update)  
DT 01-JAN-1999 (TREMBlrel. 09, Last annotation update)  
DE HYPOTHETICAL 23.3 KD PROTEIN.  
OS Synechocystis sp. (strain PCC 6803).  
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-PCC6803;

RA TABATA S.;  
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.

RC STRAIN-PCC6803;  
RX MEDLINE; 97061201.  
RA KANEKO T., SATO S., KOTANI H., TANAKA A., ASAMIZU E., NAKAMURA Y.,  
RA MIYAJIMA N., HIROSAWA M., SUGIURA M., SASAMOTO S., KIMURA T.,  
RA HOSOUCHI T., MATSUNO A., MURAKI A., NAKAZAKI N., NARUO K., OKUMURA S.,  
RA SHIMO S., TAKEUCHI C., WADA T., WATANABE A., YAMADA M., YASUDA M.,  
RA TABATA S.;  
RL "Sequence analysis of the genome of the unicellular cyanobacterium  
RT Synechocystis sp. strain PCC6803. II. Sequence determination of the  
RT entire genome and assignment of potential protein-coding regions."  
RL DNA Res. 3:109-136(1996).  
DR EMBL; D90914; BAA18445.1; -;  
KW Hypothetical protein.  
SQ SEQUENCE 211 AA; 23260 MW; 21A6D8C1 CRC32;

Query Match 100.0%; Score 19; DB 2; Length 211;  
Best Local Similarity 25.0%; Pred. No. 5.59e+00;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 203 PGDAQLLM 210  
|  
QY 2 PXXXXXXM 9

RESULT 5 PRELIMINARY; PRT; 215 AA.  
ID Q59436  
AC Q59436;  
DT 01-NOV-1996 (TREMBlrel. 01, Created)  
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)  
DE CHLORAMPHENICOL ACETYLTRANSFERASE (EC 2.3.1.28).  
GN CATPFP501.  
OS Enterococcus faecalis (Streptococcus faecalis).  
OG Plasmid PRE25.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Enterococcaceae;  
OC Enterococcus.  
RN [1]  
RP SEQUENCE FROM N.A.

RC STRAIN-RE25;  
RA PERRETEN V., MOSCHETTI G., TEUBER M.;  
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: THIS ENZYME IS AN EFFECTOR OF CHLORAMPHENICOL RESISTANCE  
CC IN BACTERIA.

CC -1- CATALYTIC ACTIVITY: ACETYL-COA + CHLORAMPHENICOL - COA +  
CC CHLORAMPHENICOL 3-ACETATE.  
CC -1- SUBUNIT: HOMOTRIMER (BY SIMILARITY).  
CC -1- SIMILARITY: HIGH, TO OTHER CAT.  
DR EMBL; X92945; CAA63498.1; -;  
DR HSSP; P00484; LOCA.  
DR PROSITE; PS00100; CAT; 1.  
DR PFAM; PF00302; CAT; 1.  
KW Transferase; Plasmid; Antibiotic resistance; Acyltransferase.  
FT ACT\_SITE 190 190 BY SIMILARITY.  
SQ SEQUENCE 215 AA; 25747 MW; AA344F22 CRC32;

Query Match 100.0%; Score 19; DB 2; Length 215;  
Best Local Similarity 25.0%; Pred. No. 5.59e+00;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 51 PSLIYAIM 58  
|  
QY 2 PXXXXXXM 9

RESULT 6 PRELIMINARY; PRT; 217 AA.  
ID O58748  
AC O58748;  
DT 01-AUG-1998 (TREMBlrel. 07, Created)  
DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)

\*\*\*\*\*

W A S E R E H  
(TM)

\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Sat Apr 15 02:19:24 2000; MasPar time 7.48 seconds  
83.471 Million cell updates/sec

Tabular output not generated.

Title: >US-08-452-843-30  
Description: (1-9) from US08452843.pep  
Perfect Score: 19  
Sequence: 1 PXXXXXXM 9

Scoring table: PAM 150  
Gap 15

Searched: 225878 seqs, 69334122 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: sptrcmbl12  
1:sp\_archaea 2:sp\_bacteria 3:sp\_fungi 4:sp\_human  
5:sp\_invertebrate 6:sp\_mammal 7:sp\_mhc 8:sp\_organelle  
9:sp\_phase 10:sp\_plant 11:sp\_rodent 12:sp\_unclassified  
13:sp\_vertebrate 14:sp\_virus

Statistics: Mean 11.476; Variance 2.544; scale 4.512

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	ID	Description	Pred. No.
1	19	100.0	107	1 P94122	HYPOTHETICAL 12.0 KD P
2	19	100.0	135	2 P77688	FROM BASES 522240 TO 5
3	19	100.0	202	2 P73219	FIBRILLIN
4	19	100.0	211	2 P74351	HYPOTHETICAL 23.3 KD P
5	19	100.0	215	2 Q59436	CHLOROPHENICOL ACETYL
6	19	100.0	217	1 Q58748	217AA LONG HYPOTHETICA
7	19	100.0	222	1 P95933	ORF C01035.
8	19	100.0	227	1 P95962	ORF C04023.
9	19	100.0	235	1 Q9VEN3	235AA LONG HYPOTHETICA
10	19	100.0	240	2 Q25044	CYTOCHROME C BIOGENESI
11	19	100.0	243	2 P73217	HYPOTHETICAL 26.6 KD P
12	19	100.0	255	2 P72952	HYPOTHETICAL 27.1 KD P
13	19	100.0	256	2 P72950	HYPOTHETICAL 28.7 KD P
14	19	100.0	261	1 Q28627	CONSERVED HYPOTHETICAL
15	19	100.0	274	2 P96189	HYPOTHETICAL 30.2 KD P
16	19	100.0	287	2 Q56563	UREASE.
17	19	100.0	294	2 Q53145	MRNA.
18	19	100.0	308	2 Q52270	LIPASE.
19	19	100.0	313	2 Q53149	ABC-TRANSPORTER ATP-BI
20	19	100.0	316	2 P94982	HYPOTHETICAL 34.0 KD P

21	19	100.0	318	2 Q48793	TMS AND PRS GENES, PAR
22	19	100.0	326	2 Q45928	QPH1 PLASMID, COMPLETE
23	19	100.0	352	2 Q9ZM20	PEPTIDE CHAIN RELEASE
24	19	100.0	365	1 Q59060	365AA LONG HYPOTHETICA
25	19	100.0	389	2 Q48816	HELIC AND ORF2 PROTEIN
26	19	100.0	391	2 Q39249	PYRUVATE KINASE (EC 2.
27	19	100.0	393	2 Q48821	PHOSPHOPENTOMUTASE (EC
28	19	100.0	396	1 Q28239	CONSERVED HYPOTHETICAL
29	19	100.0	417	1 Q30201	3-KETOACYL-COA THIOLAS
30	19	100.0	419	2 Q53861	PUTATIVE MEMBRANE TRAN
31	19	100.0	463	2 Q34944	PUTATIVE PEPTIDASE.
32	19	100.0	469	1 Q31075	ATP SYNTHASE BETA SUBU
33	19	100.0	480	2 P74370	HYPOTHETICAL 54.1 KD P
34	19	100.0	481	2 Q9ZLV5	CAG ISLAND PROTEIN.
35	19	100.0	488	2 Q51893	HYPOTHETICAL 55.9 KD P
36	19	100.0	506	1 Q53767	CYTOCHROME B.
37	19	100.0	537	2 Q85887	HYPOTHETICAL 62.2 KD P
38	19	100.0	543	2 Q53406	FATTY ACID COA-LIGASE.
39	19	100.0	564	2 Q53539	HYPOTHETICAL 59.6 KD P
40	19	100.0	636	1 Q9YF10	636AA LONG HYPOTHETICA
41	19	100.0	653	2 P77881	P-TYPE ATPASE.
42	19	100.0	675	2 Q46838	ORF_F675.
43	19	100.0	715	2 Q59208	ESTERASE (EC 3.1.1.1)
44	19	100.0	929	2 Q32491	PUTATIVE N6-ADEININE S
45	19	100.0	1048	1 Q58677	HYPOTHETICAL PROTEIN M

## ALIGNMENTS

RESULT 1  
ID P94122 PRELIMINARY; PRT; 107 AA.  
AC P94122;  
DT 01-MAY-1997 (Tremblrel. 03, Created)  
DT 01-MAY-1997 (Tremblrel. 03, Last sequence update)  
DT 01-AUG-1998 (Tremblrel. 07, Last annotation update)  
DE HYPOTHETICAL 12.0 KD PROTEIN.  
OS Acidianus ambivalens (Desulfurolobus ambivalens).  
OC Archaea; Crenarchaeota; Sulfolobales; Acidianus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-DSM 3772;  
RX MEDLINE; 97175566.  
RA PURSCHKE W.G., SCHMIDT C.L., PETERSEN A., SCHAEFER G.;  
RT "The terminal quinol oxidase of the hyperthermophilic archaeson  
RT Acidianus ambivalens exhibits a novel subunit structure and gene  
RT organization."  
RL J. Bacteriol. 179:1344-1353(1997).  
DR EMBL; Y09614; CAA70829.1; -.  
KW Hypothetical protein.  
SQ SEQUENCE 107 AA; 11955 MW; A64F49DE CRC32;

Query Match 100.0%; Score 19; DB 1; Length 107;  
Best Local Similarity 25.0%; Pred. No. 5.59e+00;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 7 PEKDSLEM 14  
Qy 2 PXXXXXXM 9

RESULT 2  
ID P77688 PRELIMINARY; PRT; 135 AA.  
AC P77688;  
DT 01-FEB-1997 (Tremblrel. 02, Created)  
DT 01-FEB-1997 (Tremblrel. 02, Last sequence update)  
DT 01-FEB-1997 (Tremblrel. 02, Last annotation update)  
DE FROM BASES 522240 TO 533123  
DE (SECTION 46 OF 400) OF THE COMPLETE GENOME (SECTION 46 OF 400).  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Escherichia.  
RN [1]  
RP SEQUENCE FROM N.A.



CC - PYROPHOSPHATE + TRANS-TRANS-FARNESYL DIPHOSPHATE.  
CC -!- CATALYTIC ACTIVITY: TRANS-TRANS-FARNESYL DIPHOSPHATE + ISOPENTENYL  
CC DIPHOSPHATE -> PYROPHOSPHATE + GERANYLGERANYL DIPHOSPHATE.  
CC -!- PATHWAY: BIOSYNTHESIS OF MEMBRANE ETHER-LINKED LIPIDS.  
CC -!- SIMILARITY: BELONGS TO THE FPP/GGPP SYNTHETASES FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL: Z77165; CAB01025.1; -  
DR PROSITE: PS00444; POLYPRENYL\_SYNTHET\_2; 1.  
DR PROSITE: PS00723; POLYPRENYL\_SYNTHET\_1; 1.  
DR PFAM: PF00348; polyprenyl\_synt; 1.  
KW Hypothetical protein; Lipid synthesis; Isoprene biosynthesis;  
KW Transferase.  
SQ SEQUENCE 359 AA; 38852 MW; 00A4138C CRC32;

Query Match 100.0%; Score 19; DB 1; Length 359;  
Best Local Similarity 25.0%; Pred. No. 2.15e+02;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 12 POPDSRDM 19  
|  
QY 2 PXXXXXXM 9

RESULT 15  
ID IL3R\_HUMAN STANDARD; PRT; 378 AA.  
AC P26951;  
DT 01-OCT-1993 (Rel. 27, Created)  
DT 01-OCT-1993 (Rel. 27, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE INTERLEUKIN-3 RECEPTOR ALPHA CHAIN PRECURSOR (IL-3R-ALPHA) (CD123  
DE ANTIGEN).  
GN IL3RA OR IL3R.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 92005668.  
RA KITAMURA T., SATO N., ARAI K., MIYAJIMA A.;  
RT "Expression cloning of the human IL-3 receptor cDNA reveals a shared  
RT beta subunit for the human IL-3 and GM-CSF receptors.";  
RL Cell 66:1165-1174(1991).  
CC -!- FUNCTION: THIS IS A RECEPTOR FOR INTERLEUKIN-3.  
CC -!- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA CHAIN WHICH IS  
CC SHARED BY THE INTERLEUKIN-3, INTERLEUKIN-5, AND GM-CSF RECEPTORS.  
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.  
CC -!- SIMILARITY: BELONGS TO THE CYTOKINE FAMILY OF RECEPTORS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL: M74782; AA59148.1; -  
DR PIR: A40266; A40266.  
DR MIM: 308385; -  
DR MIM: 430000; -  
DR PROSITE: PS00241; RECEPTOR\_CYTOKINES\_1; FALSE\_NEG.  
DR PROSITE: PS00340; RECEPTOR\_CYTOKINES\_2; FALSE\_NEG.  
KW Receptor; Transmembrane; Glycoprotein; Signal.  
FT SIGNAL 1 18 POTENTIAL.  
FT CHAIN 19 378 INTERLEUKIN-3 RECEPTOR ALPHA CHAIN.

FT DOMAIN 19 305 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 306 325 POTENTIAL.  
FT DOMAIN 326 378 CYTOPLASMIC (POTENTIAL).  
FT CARBOHYD 46 46 POTENTIAL.  
FT CARBOHYD 64 64 POTENTIAL.  
FT CARBOHYD 80 80 POTENTIAL.  
FT CARBOHYD 109 109 POTENTIAL.  
FT CARBOHYD 212 212 POTENTIAL.  
FT CARBOHYD 218 218 POTENTIAL.  
SQ SEQUENCE 378 AA; 43330 MW; 5E9F0812 CRC32;

Query Match 100.0%; Score 19; DB 1; Length 378;  
Best Local Similarity 25.0%; Pred. No. 2.15e+02;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 25 PPITNLRM 32  
|  
QY 2 PXXXXXXM 9

Search completed: Sat Apr 15 02:19:06 2000  
Job time : 42 secs.

DE GLYCERALDEHYDE 3-PHOSPHATE DEHYDROGENASE, CYTOSOLIC (EC 1.2.1.12).

GN GAPC.

OS Physcomitrella patens (Moss).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Bryopsida;

OC Bryidae; Funariales; Funariaceae; Physcomitrella.

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE; 93196483.

RA MARTIN W., LYDIATE D., BRINKMANN H., FORKMANN G., SAEDLER H.,

RA CERFF R.;

RT "Molecular phylogenies in angiosperm evolution.";

RL Mol. Biol. Evol. 10:140-162(1993).

CC -1- CATALYTIC ACTIVITY: D-GLYCERALDEHYDE 3-PHOSPHATE + ORTHOPHOSPHATE

CC + NAD(+) = 1,3-DIPHOSPHATEGLYCERATE + NADH.

CC -1- PATHWAY: FIRST STEP IN THE SECOND PHASE OF GLYCOLYSIS.

CC -1- SUBUNIT: HOMOTETRAMER.

CC -1- SUBCELLULAR LOCATION: CYTOSOL.

CC -1- SIMILARITY: BELONGS TO THE GLYCERALDEHYDE 3-PHOSPHATE

CC DEHYDROGENASE FAMILY.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC

DR EMBL; X72381; CAA51071.1; -

DR HSP; P00357; LGPD. -

DR MENDEL; 568; Physa; GapC; 1.

DR PROSITE; PS00071; GAPDH; 1.

DR PFAM; PF00044; gpdh; 1.

KW Glycolysis; Oxidoreductase; NAD; Multigene family.

FT BINDING 156 156 GLYCERALDEHYDE 3-PHOSPHATE.

FT ACT SITE 183 183 ACTIVATES THIOL GROUP DURING CATALYSIS.

FT SEQUENCE 342 AA; 36762 MW; 64412D47 CRC32;

Query Match 100.0%; Score 19; DB 1; Length 342;

Best Local Similarity 25.0%; Pred. No. 2.15e+02;

Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 39 PFITPEYM 46

QY 2 PXXXXXXM 9

RESULT 13

ID GB0\_LYMET STANDARD; PRT; 353 AA.

AC P30683;

DT 01-APR-1993 (Rel. 25, Created)

DT 01-OCT-1994 (Rel. 30, Last sequence update)

DT 01-OCT-1996 (Rel. 34, Last annotation update)

DE GUANINE NUCLEOTIDE-BINDING PROTEIN G(O), ALPHA SUBUNIT.

OS Lymnaea stagnalis (Great pond snail).

OC Eukaryota; Metazoa; Mollusca; Gastropoda; Pulmonata; Basommatophora;

OC Lymnaeidae; Lymnaea.

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE-CNS;

RX MEDLINE; 93106153.

RA KNOL J.C., WEIDEMANN W., PLANTA R.J., VREUGDENHIL E.,

RA VAN HEERIKUIZEN H.;

RT "Molecular cloning of G protein alpha subunits from the central

RT nervous system of the mollusc Lymnaea stagnalis.";

RL FEBS Lett. 314:215-219(1992).

CC -1- FUNCTION: GUANINE NUCLEOTIDE-BINDING PROTEINS (G PROTEINS) ARE

CC INVOLVED AS MODULATORS OR TRANSDUCERS IN VARIOUS TRANSMEMBRANE

CC SIGNALING SYSTEMS.

CC -1- FUNCTION: THE G(O) PROTEIN FUNCTION IS NOT CLEAR.

CC -1- SUBUNIT: G PROTEINS ARE COMPOSED OF 3 UNITS (ALPHA, BETA & GAMMA).

CC THE ALPHA CHAIN CONTAINS THE GUANINE NUCLEOTIDE BINDING SITE.

CC -1- SIMILARITY: BELONGS TO THE G-ALPHA FAMILY. SUBFAMILY 1

CC (G(I/O/T/2)).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC

DR EMBL; Z15094; CAA78806.1; -

DR PIR; S25589; S25589.

DR PIR; S27014; S27014.

DR HSP; P04896; IAST.

KW GTP-binding; Transducer; Multigene family; ADP-ribosylation;

KW Myristate; Palmitate; Lipoprotein.

FT INIT\_MET 0 0 BY SIMILARITY.

FT LIPID 1 1 MYRISTATE (BY SIMILARITY).

FT LIPID 2 2 PALMITATE (BY SIMILARITY).

FT NP\_BIND 39 46 GTP (BY SIMILARITY).

FT NP\_BIND 200 203 GTP (BY SIMILARITY).

FT NP\_BIND 269 272 GTP (BY SIMILARITY).

FT MOD\_RES 178 178 ADP-RIBOSYL[1] (BY ACTION OF CTX)

FT MOD\_RES 350 350 (BY SIMILARITY).

FT MOD\_RES 350 350 ADP-RIBOSYL[1] (BY ACTION OF IAP).

FT SEQUENCE 353 AA; 40140 MW; ECB09F4A CRC32;

Query Match 100.0%; Score 19; DB 1; Length 353;

Best Local Similarity 25.0%; Pred. No. 2.15e+02;

Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 288 PEYTGKQM 295

QY 2 PXXXXXXM 9

RESULT 14

ID GGPP\_MYCTU STANDARD; PRT; 359 AA.

AC Q50727;

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 15-DEC-1999 (Rel. 39, Last annotation update)

DE PROBABLE GERANYLGERANYL PYROPHOSPHATE SYNTHETASE (GGPP SYNTHETASE)

DE [INCLUDES: DIMETHYLLALLYLTRANSFERASE (EC 2.5.1.1);

DE GERANYLTRANSFERASE (EC 2.5.1.10); FARNESYLTRANSFERASE

DE (EC 2.5.1.29)]

GN RV3398C OR MTCY78.30.

OS Mycobacterium tuberculosis.

OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;

OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-H37RV;

RX MEDLINE; 98295987.

RA COLE S.T., BROSCH R., PARKHILL J., GARNIER T., CHURCHER C., HARRIS D.,

RA GORDON S.V., EIGLMEIER K., GAS S., BARRY C.E. III, TEKAI F.,

RA BADCOCK K., BASHAM D., BROWN D., CHILLINGWORTH T., CONNOR R.,

RA DAVIES R., DEVLIN K., FELTWEILL T., GENTLES S., HAMLIN N., HOLROYD S.,

RA HORNSBY T., JAGELS K., KROGH A., MCLEAN J., MOULE S., MURPHY L.,

RA OLIVER S., OSBORNE J., QUAIL M.A., RAJANDREAM M.A., ROGERS J.,

RA RUTTER S., SEEGER K., SKELTON S., SQUARES S., SQUARES R., SULSTON J.E.,

RA TAYLOR K., WHITEHEAD S., BARRELL B.G.;

RT "Deciphering the biology of Mycobacterium tuberculosis from the

RT complete genome sequence.";

RL Nature 393:537-544(1998).

CC -1- FUNCTION: CATALYZES THE TRANS-ADDITION OF THE THREE MOLECULES OF

CC IPP ONTO DMAPP TO FORM GERANYLGERANYL PYROPHOSPHATE WHICH IS A

CC PRECURSOR OF THE ETHER-LINKED LIPIDS.

CC -1- CATALYTIC ACTIVITY: DIMETHYLLALLYL DIPHOSPHATE + ISOPENTENYL

CC DIPHOSPHATE = PYROPHOSPHATE + GERANYL DIPHOSPHATE

CC -1- CATALYTIC ACTIVITY: GERANYL DIPHOSPHATE + ISOPENTENYL DIPHOSPHATE

CC FORM MPF.  
CC -1- SIMILARITY: BELONGS TO THE CYCLIN FAMILY. CYCLIN C SUBFAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; D14013; BAA03114.1; -  
DR HSP; P51946; IJWK.  
DR PROSITE; PS00292; CYCLINS; FALSE NEG.  
KW Cyclin; Cell cycle; Cell division.  
SQ SEQUENCE 298 AA; 34845 MW; F0748653 CRC32;  
  
Query Match 100.0%; Score 19; DB 1; Length 298;  
Best Local Similarity 25.0%; Pred. No. 2.15e+02;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
  
Db 140 PKEPPYEM 147  
QY 2 PXXXXXXM 9  
  
RESULT 10  
ID CATV\_NPYEM STANDARD; PRT; 323 AA.  
AC P41721;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 01-NOV-1995 (Rel. 32, Last annotation update)  
DE VIRAL CATHESIN (EC 3.4.22.-) (V-CATH).  
GN VCATH.  
OS Bombyx mori nuclear polyhedrosis virus (BnNPV).  
OC Viruses; dsDNA viruses, no RNA stage; Baculoviridae;  
OC Nucleopolyhedrovirus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-73;  
RX MEDLINE; 94365965.  
RA OHKAWA T., MAJIMA K., MAEDA S.;  
RT "A cysteine protease encoded by the baculovirus Bombyx mori nuclear  
RT polyhedrosis virus."  
RL J. Virol. 68:6619-6625(1994).  
CC -1- FUNCTION: MAY PLAY A ROLE IN DEGRADATION OF INFECTED LARVAE TO  
CC FACILITATE HORIZONTAL TRANSMISSION OF THE VIRUS.  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C1; ALSO KNOWN AS THE  
CC PAPAIN FAMILY OF THIOL PROTEASES.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; U12688; AAB49542.1; -  
DR HSP; P43235; IATK.  
DR PROSITE; PS00139; THIOL PROTEASE CYS; 1.  
DR PROSITE; PS00639; THIOL PROTEASE HIS; 1.  
DR PROSITE; PS00640; THIOL PROTEASE ASN; 1.  
DR PFAM; PF00112; Peptidase\_C1; 1.  
KW Hydrolyase; Thiol protease.  
FT ACT\_SITE 136 136 BY SIMILARITY.  
FT ACT\_SITE 269 269 BY SIMILARITY.  
FT ACT\_SITE 289 289 BY SIMILARITY.  
FT DISULFID 133 174 BY SIMILARITY.  
FT DISULFID 167 207 BY SIMILARITY.  
FT DISULFID 262 310 BY SIMILARITY.  
SQ SEQUENCE 323 AA; 36922 MW; A3E34B9E CRC32;

Query Match 100.0%; Score 19; DB 1; Length 323;  
Best Local Similarity 25.0%; Pred. No. 2.15e+02;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
  
Db 237 PLVGPIPM 244  
QY 2 PXXXXXXM 9  
  
RESULT 11  
ID G3PC\_PETHY STANDARD; PRT; 337 AA.  
AC P26520;  
DT 01-AUG-1992 (Rel. 23, Created)  
DT 01-AUG-1992 (Rel. 23, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE GLYCERALDEHYDE 3-PHOSPHATE DEHYDROGENASE, CYTOSOLIC (EC 1.2.1.12).  
GN GAPC OR GAPDH.  
OS Petunia hybrida (Petunia).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;  
OC core eudicots; Asteridae; euasterids I; Solanales; Solanaceae;  
OC Petunia.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA MARTIN W., GIERL A., SAEDLER H.;  
RT "Molecular evidence for pre-Cretaceous angiosperm origins."  
RL Nature 339:46-48(1989).  
CC -1- CATALYTIC ACTIVITY: D-GLYCERALDEHYDE 3-PHOSPHATE + ORTHOPHOSPHATE  
CC + NAD(+) = 1,3-DIPHOSPHATEGLYCERATE + NADH.  
CC -1- PATHWAY: FIRST STEP IN THE SECOND PHASE OF GLYCOLYSIS.  
CC -1- SUBUNIT: HOMOTETRAMER.  
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.  
CC -1- MISCELLANEOUS: PLANTS CONTAIN THREE FORMS OF GAPDH: A CYTOSOLIC  
CC FORM WHICH PARTICIPATES IN GLYCOLYSIS AND TWO CHLOROPLAST FORMS  
CC WHICH PARTICIPATES IN PHOTOSYNTHESIS. THESE THREE FORMS ARE  
CC ENCODED BY DISTINCT GENES.  
CC -1- SIMILARITY: BELONGS TO THE GLYCERALDEHYDE 3-PHOSPHATE  
CC DEHYDROGENASE FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; X60346; CAA42904.1; -  
DR PIR; S18485; DEPJG.  
DR HSP; P00357; IGPD.  
DR MENDEL; 375; PETHY; GapC; 1.  
DR PROSITE; PS00071; GAPDH; 1.  
DR PFAM; PF00044; gpdh; 1.  
KW Glycolysis; Oxidoreductase; NAD; Multigene family.  
FT BINDING 154 154 GLYCERALDEHYDE 3-PHOSPHATE.  
FT ACT\_SITE 181 181 ACTIVATES THIOL GROUP DURING CATALYSIS.  
SQ SEQUENCE 337 AA; 36527 MW; 269BC80E CRC32;  
  
Query Match 100.0%; Score 19; DB 1; Length 337;  
Best Local Similarity 25.0%; Pred. No. 2.15e+02;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
  
Db 36 PFISVEYM 43  
QY 2 PXXXXXXM 9  
  
RESULT 12  
ID G3PC\_PHYPA STANDARD; PRT; 342 AA.  
AC P34923;  
DT 01-FEB-1994 (Rel. 28, Created)  
DT 01-FEB-1994 (Rel. 28, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)



01-DEC-1992 (Rel. 24, Last sequence update)  
15-JUL-1999 (Rel. 38, Last annotation update)  
CYTOCHROME B6.  
PETB.  
Chlamydomonas reinhardtii.  
Chloroplast.  
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;  
Chlamydomonadaceae; Chlamydomonas.  
[1]  
SEQUENCE FROM N.A.  
MEDLINE: 92256821.  
RA HUANG C., LIU X.-Q.;  
"Nucleotide sequence of the frxC, petB and trnL genes in the  
chloroplast genome of Chlamydomonas reinhardtii.";  
Plant Mol. Biol. 18:985-988(1992).  
[2]  
SEQUENCE FROM N.A.  
STRAIN-137C;  
MEDLINE: 91285146.  
RA BUESCHLEN S., CHOQUET Y., KURAS R., WOLLMAN F.A.;  
"Nucleotide sequences of the continuous and separated petA, petB and  
petD chloroplast genes in Chlamydomonas reinhardtii.";  
FEBS Lett. 284:257-262(1991).  
[3]  
CHARACTERIZATION.  
RC STRAIN-WT12;  
RA PIERRE Y., BREYTON C., KRAMER D., POPOUT J.-L.;  
"Purification and characterization of the cytochrome b6 f complex  
from Chlamydomonas reinhardtii.";  
J. Biol. Chem. 270:29342-29349(1995).  
CC -!- FUNCTION: COMPONENT OF THE CYTOCHROME B6/F COMPLEX WHICH IS PART  
OF THE CHLOROPLASTIC RESPIRATORY CHAIN.  
CC -!- CATALYTIC ACTIVITY: PLASTOQUINOL-1 + 2 OXYDIZED PLASTOCYANIN -  
PLASTOQUINONE + 2 REDUCED PLASTOCYANIN.  
CC -!- COFACTOR: TWO HEME GROUPS (B562 AND B566) WHICH ARE NOT COVALENTLY  
BOUND TO THE PROTEIN.  
CC -!- SUBUNIT: THE MAIN SUBUNITS OF COMPLEX B6-F ARE: CYTOCHROME B6,  
17 KD POLYPEPTIDE (PETD), CYTOCHROME F AND THE RIESKE PROTEIN.  
CC -!- PTM: THE N-TERMINUS IS BLOCKED.  
CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME B FAMILY. CORRESPONDS TO THE  
AMINO END OF MITOCHONDRIAL CYTOCHROME B.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DBL: X62905; CAA44690.1; -;  
DR EMBL: X72918; CAA51423.1; -;  
DR PIR: S16917; S16917.  
DR PIR: S21253; S21253.  
DR MENDEL: 4387; CHLre; petB.1.  
DR PROSITE: P500192; CYTOCHROME\_B\_HEME; 1.  
DR PFAM: PF00033; cytochrome\_b\_n; 1.  
KW Electron transport; Heme; Chloroplast; Photosynthesis; Transmembrane.  
FT METAL 86 86  
IRON 1 (HEME AXIAL LIGAND)  
(BY SIMILARITY).  
FT METAL 100 100  
IRON 2 (HEME AXIAL LIGAND)  
(BY SIMILARITY).  
FT METAL 187 187  
IRON 2 (HEME AXIAL LIGAND)  
(BY SIMILARITY).  
FT METAL 202 202  
IRON 1 (HEME AXIAL LIGAND)  
(BY SIMILARITY).  
FT SEQUENCE 215 AA; 24164 MW; E18A90C1 CRC32;  
Query Match 100.0%; Score 19; DB 1; Length 215;  
Best Local Similarity 25.0%; Pred.No. 2.15e+02;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
-----

DB 192 PLLTAVFM 199  
QY 2 PXXXXXXM 9  
RESULT 8  
ID BRUL\_SOYBN STANDARD; PRT; 283 AA.  
AC P35694;  
DT 01-JUN-1994 (Rel. 29, Created)  
DT 01-JUN-1994 (Rel. 29, Last sequence update)  
DT 01-OCT-1996 (Rel. 34, Last annotation update)  
DE BRASSINOSTEROID-REGULATED PROTEIN BRUL.  
OS Glycine max (Soybean).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC eukaryotes; Spermatophyta; Magnoliophyta; eudicotyledons;  
OC core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae;  
OC Glycine.  
CC [1]  
SEQUENCE FROM N.A.  
RC TISSUE-EPICOTYL;  
RX MEDLINE: 94159788.  
RA ZUREK D.M., CLOUSE S.D.;  
"Molecular cloning and characterization of a  
brassinosteroid-regulated gene from elongating soybean (Glycine max  
L.) epicotyls.";  
Plant Physiol. 104:161-170(1994).  
CC -!- FUNCTION: POSSIBLE ROLE IN BRASSINOSTEROID-STIMULATED ELONGATION.  
CC -!- SIMILARITY: TO A THALIANA MERI-5.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DBL: L22162; AAA81350.1; -;  
DR HSP: P23904; LAJK.  
DR PFAM: PF00722; Glyco\_hydro\_16; 1.  
SQ SEQUENCE 283 AA; 32254 MW; 2523BB0C CRC32;  
Query Match 100.0%; Score 19; DB 1; Length 283;  
Best Local Similarity 25.0%; Pred.No. 2.15e+02;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
-----  
DB 181 PFPKQPM 188  
QY 2 PXXXXXXM 9  
RESULT 9  
ID CGIC-RAT STANDARD; PRT; 298 AA.  
AC P39947;  
DT 01-FEB-1995 (Rel. 31, Created)  
DT 01-FEB-1995 (Rel. 31, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE GLYS-SPECIFIC CYCLIN C.  
GN CCNC.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
CC [1]  
SEQUENCE FROM N.A.  
RC TISSUE-KIDNEY;  
RX MEDLINE: 9330551.  
RA TAMURA K., KANAOKA Y., JINNO S., NAGATA A., OGISO Y., SHIMIZU K.,  
RA HAYAKAWA T., NOJIMA H., OKAYAMA H.;  
"Cyclin G: a new mammalian cyclin with homology to fission yeast  
Cig1.";  
Oncogene 8:2113-2118(1993).  
RL Oncogene 8:2113-2118(1993).  
CC -!- FUNCTION: ESSENTIAL FOR THE CONTROL OF THE CELL CYCLE AT THE G1/S  
(START) TRANSITION. INTERACTS WITH THE CDC2 PROTEIN KINASE TO

```
CC -1- SIMILARITY: STRONG, TO OTHER P20-ARC SUBUNITS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X74152; -; NOT_ANNOTATED.CDS.
DR EMBL; S53418; AAB24905.1; ALT_TERM.
DR EMBL; Z28013; CAA81848.1; -
DR PIR; S37826; S37826.
DR SGD; L0004034; ARC19.
SQ SEQUENCE 171 AA; 19916 MW; FAALCFSE CRC32;
Query Match 100.0%; Score 19; DB 1; Length 171;
Best Local Similarity 25.0%; Pred. No. 2.15e+02;
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
Db 66 PSVNSVRM 73
|
QY 2 PXXXXXXM 9
RESULT 5
ID GLPF_MYCGA STANDARD; PRT; 205 AA.
AC P52280;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE GLYCEROL UPTAKE FACILITATOR PROTEIN.
GN GLPF.
OS Mycoplasma gallisepticum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC Mycoplasmataceae; Mycoplasma.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=56;
RA FORSYTH M.H., SAUD S., GEARY S.J.;
RL Submitted (OCT-1995) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: GLYCEROL ENTERS THE CELL VIA THE GLYCEROL DIFFUSION
CC FACILITATOR PROTEIN. THIS MEMBRANE PROTEIN FACILITATES THE
CC MOVEMENT OF GLYCEROL ACROSS THE CYTOPLASMIC MEMBRANE (BY
CC SIMILARITY).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE TRANSMEMBRANE CHANNEL MIP FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U35010; AAA79047.1; -
DR PROSITE; PS00221; MIP; 1.
DR PFAM; PF00230; MIP; 2.
KW Glycerol metabolism; Transport; Transmembrane.
FT TRANSMEM 7 27 POTENTIAL.
FT TRANSMEM 45 65 POTENTIAL.
FT TRANSMEM 88 108 POTENTIAL.
FT TRANSMEM 143 163 POTENTIAL.
FT TRANSMEM 166 186 POTENTIAL.
SQ SEQUENCE 205 AA; 22225 MW; 626630AA CRC32;
Query Match 100.0%; Score 19; DB 1; Length 205;
Best Local Similarity 25.0%; Pred. No. 2.15e+02;
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
Db 73 PAVTVTFM 80
-----
CC -1- SIMILARITY: STRONG, TO OTHER P20-ARC SUBUNITS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X74152; -; NOT_ANNOTATED.CDS.
DR EMBL; S53418; AAB24905.1; ALT_TERM.
DR EMBL; Z28013; CAA81848.1; -
DR PIR; S37826; S37826.
DR SGD; L0004034; ARC19.
SQ SEQUENCE 171 AA; 19916 MW; FAALCFSE CRC32;
Query Match 100.0%; Score 19; DB 1; Length 171;
Best Local Similarity 25.0%; Pred. No. 2.15e+02;
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
Db 66 PSVNSVRM 73
|
QY 2 PXXXXXXM 9
RESULT 5
ID GLPF_MYCGA STANDARD; PRT; 205 AA.
AC P52280;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE GLYCEROL UPTAKE FACILITATOR PROTEIN.
GN GLPF.
OS Mycoplasma gallisepticum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC Mycoplasmataceae; Mycoplasma.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=56;
RA FORSYTH M.H., SAUD S., GEARY S.J.;
RL Submitted (OCT-1995) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: GLYCEROL ENTERS THE CELL VIA THE GLYCEROL DIFFUSION
CC FACILITATOR PROTEIN. THIS MEMBRANE PROTEIN FACILITATES THE
CC MOVEMENT OF GLYCEROL ACROSS THE CYTOPLASMIC MEMBRANE (BY
CC SIMILARITY).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE TRANSMEMBRANE CHANNEL MIP FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U35010; AAA79047.1; -
DR PROSITE; PS00221; MIP; 1.
DR PFAM; PF00230; MIP; 2.
KW Glycerol metabolism; Transport; Transmembrane.
FT TRANSMEM 7 27 POTENTIAL.
FT TRANSMEM 45 65 POTENTIAL.
FT TRANSMEM 88 108 POTENTIAL.
FT TRANSMEM 143 163 POTENTIAL.
FT TRANSMEM 166 186 POTENTIAL.
SQ SEQUENCE 205 AA; 22225 MW; 626630AA CRC32;
Query Match 100.0%; Score 19; DB 1; Length 205;
Best Local Similarity 25.0%; Pred. No. 2.15e+02;
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
Db 73 PAVTVTFM 80
-----
QY 2 PXXXXXXM 9
RESULT 6
ID CYB6_CHLPR STANDARD; PRT; 215 AA.
AC P13347;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE CYTOCHROME B6.
GN PETB.
OS Chlorella protothecoides.
OG Chloroplast.
OC Eukaryota; Viridiplantae; Chlorophyta; Trebouxiophyceae; Chlorellales;
OC Chlorellaceae; Chlorella.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=211-7A;
RX MEDLINE; 92003673.
RA REIMANN A., KUECK U.;
RL 'Nucleotide sequence of the plastid genes for apocytochrome b6 (petb)
RL and subunit IV of the cytochrome b6-f complex (petd) from the green
RL alga Chlorella protothecoides: lack of introns.';
RL Plant Mol. Biol. 13:255-256(1989).
CC -1- FUNCTION: COMPONENT OF THE CYTOCHROME B6/F COMPLEX WHICH IS PART
CC OF THE CHLOROPLASTIC RESPIRATORY CHAIN.
CC -1- CATALYTIC ACTIVITY: PLASTOQUINOL-1 + 2 OXYDIZED PLASTOCYANIN -
CC PLASTOQUINONE + 2 REDUCED PLASTOCYANIN.
CC -1- COFACTOR: TWO HEME GROUPS (B562 AND B566) WHICH ARE NOT COVALENTLY
CC BOUND TO THE PROTEIN.
CC -1- SUBUNIT: THE MAIN SUBUNITS OF COMPLEX B6-F ARE: CYTOCHROME B6,
CC 17 KD POLYPEPTIDE (PETD), CYTOCHROME F AND THE RIESKE PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME B FAMILY. CORRESPONDS TO THE
CC AMINO END OF MITOCHONDRIAL CYTOCHROME B.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X15244; CAA33322.1; -
DR PIR; S06159; CEK16P.
DR MENDEL; I1630; CHLPR:petb;1.
DR PROSITE; PS00192; CYTOCHROME_B_HEME; 1.
DR PFAM; PF00033; Cytochrome_b_N; 1.
KW Electron transport; Heme; Chloroplast; Photosynthesis; Transmembrane.
FT METAL 86 86 IRON 1 (HEME AXIAL LIGAND)
FT METAL 100 100 IRON 2 (HEME AXIAL LIGAND)
FT METAL 187 187 IRON 2 (HEME AXIAL LIGAND)
FT METAL 202 202 IRON 1 (HEME AXIAL LIGAND)
FT METAL 202 202 IRON 1 (HEME AXIAL LIGAND)
SQ SEQUENCE 215 AA; 24429 MW; 923CE4CD CRC32;
Query Match 100.0%; Score 19; DB 1; Length 215;
Best Local Similarity 25.0%; Pred. No. 2.15e+02;
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
Db 192 PLFTAVFM 199
|
QY 2 PXXXXXXM 9
RESULT 7
ID CYB6_CHLRE STANDARD; PRT; 215 AA.
AC Q00471;
DT 01-DEC-1992 (Rel. 24, Created)
```

DB 36 PYREVLGM 43  
QY 2 PXXXXXXM 9

RESULT 2  
ID ISB\_SHISO STANDARD; PRT; 131 AA.  
AC P19766;  
DT 01-FEB-1991 (Rel. 17, Created)  
DT 01-FEB-1991 (Rel. 17, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE INSERTION ELEMENT ISI PROTEIN INSB.  
GN INSB.  
OS Shigella sonnei.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Shigella.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA OHTSUBO E., OHTSUBO H., DOROSZKIEWICZ W., NYMAN K., ALLEN D.,  
RA DAVISON D.;  
RT "An evolutionary analysis of iso-ISI elements from Escherichia coli  
and Shigella strains."  
RL J. Gen. Appl. Microbiol. 30:359-376(1984).  
CC -|- FUNCTION: ABSOLUTELY REQUIRED FOR TRANSPOSITION OF ISI.  
CC -|- SIMILARITY: TO INSB PROTEINS OF OTHER ISI ELEMENTS.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; M37615; AAA96694.1;  
DR Transposition; Transposable element; DNA recombination.  
KW SEQUENCE 131 AA; 15508 MW; 70BE25C3 CRC32;  
SQ

Query Match 100.0%; Score 19; DB 1; Length 131;  
Best Local Similarity 25.0%; Pred. No. 2.15e+02;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

DB 56 PFDVIVM 63  
QY 2 PXXXXXXM 9

RESULT 3  
ID CHEY\_TREPA STANDARD; PRT; 144 AA.  
AC P96126;  
DT 15-DEC-1998 (Rel. 37, Created)  
DT 15-DEC-1998 (Rel. 37, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE CHEMOTAXIS PROTEIN CHEY.  
GN CHEY OR TP0366.  
OS Treponema pallidum.  
OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-NICHOLS;  
RX MEDLINE; 97399391.  
RA GREENE S.R., STAMM L.V., HARDHAM J.M., YOUNG N.R., FRYE J.G.;  
RT "Identification, sequences, and expression of Treponema pallidum  
chemotaxis genes."  
RL DNA Seq. 7:267-284(1997).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-NICHOLS;  
RX MEDLINE; 98332770.  
RA FRASER C.M., NORRIS S.J., WEINSTOCK G.M., WHITE O., SUTTON G.G.,  
RA DODSON R., GWINN M., HICKEY E.K., CLAYTON R., KETCHUM K.A.,  
RA SODERGREN E., HARDHAM J.M., MCLEOD M.P., SALZBERG S., PETERSON J.,  
RA KHALAK H., RICHARDSON D., HOWELL J.K., CHIDAMBARAM M., UTTERBACK T.,

RA MCDONALD L., ARTIACH P., BOWMAN C., COTTON M.D., FUJII C., GARLAND S.,  
RA HATCH B., HORST K., ROBERTS K., WATTHEY L., WEIDMAN J., SMITH H.O.,  
RA VENTER J.C.;  
RT "Complete genome sequence of Treponema pallidum, the syphilis  
spirochete."  
RL Science 281:375-388(1998).  
CC -|- FUNCTION: INVOLVED IN THE TRANSMISSION OF SENSORY SIGNALS FROM  
CC THE CHEMORECEPTORS TO THE FLAGELLAR MOTORS. CHEY SEEMS TO REGULATE  
CC THE CLOCKWISE (CW) ROTATION (BY SIMILARITY).  
CC -|- SUBCELLULAR LOCATION: CYTOPLASMIC.  
CC -|- PTM: PHOSPHORYLATED BY CHEA (BY SIMILARITY).  
CC -|- SIMILARITY: TO REGULATORY COMPONENTS OF SENSORY TRANSDUCTION  
CC SYSTEMS.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; U61851; AAC45558.1;  
DR EMBL; AE001215; AAC85351.1;  
DR TIGR; TP0366;  
DR PFAM; PF00072; response\_reg; 1.  
KW Chemotaxis; Sensory transduction; Phosphorylation; Flagellar rotation.  
FT DOMAIN 25 144 RECEIVER DOMAIN (POTENTIAL).  
FT MOD\_RES 78 78 PHOSPHORYLATION (BY SIMILARITY).  
SQ SEQUENCE 144 AA; 15735 MW; FDBE01C8 CRC32;  
Query Match 100.0%; Score 19; DB 1; Length 144;  
Best Local Similarity 25.0%; Pred. No. 2.15e+02;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

DB 70 PGVDLVTM 77  
QY 2 PXXXXXXM 9

RESULT 4  
ID AR20\_YEAST STANDARD; PRT; 171 AA.  
AC P33204;  
DT 01-FEB-1994 (Rel. 28, Created)  
DT 01-JUN-1994 (Rel. 29, Last sequence update)  
DT 15-JUL-1998 (Rel. 36; Last annotation update)  
DE ARP2/3 COMPLEX 20 KD SUBUNIT (P20-ARC).  
GN ARC19 OR YKL013C OR YKL166.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;  
OC Saccharomycetaceae; Saccharomyces.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 94205264.  
RA WIEMANN S., VOSS H., SCHWAGER C., RUPP T., STEGEMANN J.,  
RA ZIMMERMANN J., GROTHUES D., SENSEN C., ERLE H., HEWITT N.,  
RA BANREVI A., ANSORGE W.;  
RT "Sequencing and analysis of 51.6 kilobases on the left arm of  
RT chromosome XI from Saccharomyces cerevisiae reveals 23 open reading  
RT frames including the FAS1 gene."  
RL Yeast 9:1343-1348(1993).  
RN [2]  
RP SEQUENCE OF 1-130 FROM N.A.  
RX MEDLINE; 93127732.  
RA PASCOLO S., GHAZVINI M., BOYER J., COLLEAUX L., THIERRY A., DUJON B.;  
RT "The sequence of a 9.3 kb segment located on the left arm of the  
RT yeast chromosome XI reveals five open reading frames including the  
RT CCE1 gene and putative products related to MIO2 and to the ribosomal  
RT protein L10."  
RL Yeast 8:987-995(1992).  
CC -|- FUNCTION: PART OF A COMPLEX IMPLICATED IN THE CONTROL OF ACTIN  
CC POLYMERIZATION IN CELLS (BY SIMILARITY).  
CC -|- SUBUNIT: BELONGS TO A COMPLEX COMPOSED OF ARP2, ARP3, P41-ARC,  
CC P34-ARC, P21-ARC, P20-ARC AND P16-ARC (BY SIMILARITY).

\*\*\*\*\*

W P E L E H (TM)

\*\*\*\*\*

Release 3.1A John F. Collins, BioComputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

Mpsrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Sat Apr 15 02:18:24 2000; MasPar time 3.12 Seconds  
Tabular output not generated. 86.254 Million cell updates/sec.

Title: >US-08-452-843-30  
Description: (1-9) from US08452843.pap  
Perfect Score: 19  
Sequence: 1 XPXXXXXXM 9

Scoring table: PAM 150  
Gap 15

Searched: 82229 seqs, 29864866 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: swiss-prot38  
1:swissprot

Statistics: Mean 12.133; Variance 4.890; scale 2.481

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	19	100.0	103	1	CHLB_OSMCL PROTOCHLOROPHYLLIDE RE	2.15e+02
2	19	100.0	131	1	ISB_SHISO INSERTION ELEMENT IS1	2.15e+02
3	19	100.0	144	1	CHEY_TREPA CHEMOTAXIS PROTEIN CHE	2.15e+02
4	19	100.0	171	1	AR20_YEAST ARP2/3 COMPLEX 20 KD S	2.15e+02
5	19	100.0	205	1	GLPF_MYCGA GLYCEROL UPTAKE FACILI	2.15e+02
6	19	100.0	215	1	CYB6_CHLPR CYTOCHROME B6.	2.15e+02
7	19	100.0	215	1	CYB6_CHLRE CYTOCHROME B6.	2.15e+02
8	19	100.0	283	1	BRUL_SOYBN BRASSINOSTEROID-REGULA	2.15e+02
9	19	100.0	298	1	CGIC_RAT GI/S-SPECIFIC CYCLIN C	2.15e+02
10	19	100.0	323	1	CATV_NPVBM VIRAL CATHEPSIN (EC 3.	2.15e+02
11	19	100.0	337	1	G3PC_PETHY GLYCERALDEHYDE 3-PHOSP	2.15e+02
12	19	100.0	342	1	G3PC_PHPYA GLYCERALDEHYDE 3-PHOSP	2.15e+02
13	19	100.0	353	1	G80_LYMSI GUANINE NUCLEOTIDE-BIN	2.15e+02
14	19	100.0	359	1	GGPP_MYCTU PROBABLE GERANYLGERANY	2.15e+02
15	19	100.0	378	1	IL3R_HUMAN INTERLEUKIN-3 RECEPTOR	2.15e+02
16	19	100.0	379	1	AMPC_MORMO BETA-LACTAMASE PRECURS	2.15e+02
17	19	100.0	384	1	GALI_ACTPL GALACTOKINASE (EC 2.7.	2.15e+02
18	19	100.0	389	1	CHSA_PEA CHALCONE SYNTHASE 1A (	2.15e+02
19	19	100.0	389	1	CHSA_PETHY CHALCONE SYNTHASE A (E	2.15e+02
20	19	100.0	411	1	IHH_MOUSE INDIAN HEDGEHOG PROTEI	2.15e+02
21	19	100.0	446	1	AK_RICPR ASPARTOKINASE (EC 2.7.	2.15e+02
22	19	100.0	458	1	BPHA_PSEPS BIPHENYL DIOXYGENASE A	2.15e+02
23	19	100.0	468	1	CYCA_ACEPO ALCOHOL DEHYDROGENASE	2.15e+02

24	19	100.0	471	1	5H2A_MOUSE 5-HYDROXYTRYPTAMINE 2A	2.15e+02
25	19	100.0	471	1	5H2A_RAT 5-HYDROXYTRYPTAMINE 2A	2.15e+02
26	19	100.0	476	1	CBPH_MOUSE CARBOXYPEPTIDASE H PRE	2.15e+02
27	19	100.0	494	1	CD5_MOUSE T-CELL SURFACE GLYCOPR	2.15e+02
28	19	100.0	495	1	CD5_HUMAN T-CELL SURFACE GLYCOPR	2.15e+02
29	19	100.0	507	1	ARSA_HUMAN ARYLSULFATASE A PRECUR	2.15e+02
30	19	100.0	509	1	CK56_CHICK GAP JUNCTION CX56 PROT	2.15e+02
31	19	100.0	513	1	CHLB_MARPO PROTOCHLOROPHYLLIDE RE	2.15e+02
32	19	100.0	518	1	CYTOCHROME P450 1A1 (E	2.15e+02
33	19	100.0	521	1	CP11_RABIT CYTOCHROME P450 1A1 (E	2.15e+02
34	19	100.0	538	1	CP11_PLEPL RNA EXPORT FACTOR GLE1	2.15e+02
35	19	100.0	542	1	AROF_YEAST RHO PHOSFO-2-DEHYDRO-3-DE	2.15e+02
36	19	100.0	580	1	IOLD_BACSU IOLD PROTEIN.	2.15e+02
37	19	100.0	676	1	AWYL_ECOLI ALPHA-AMYLASE PRECURSO	2.15e+02
38	19	100.0	734	1	GLGB_AGRU 1,4-ALPHA-GLUCAN BRANC	2.15e+02
39	19	100.0	902	1	ITHL_PIG INTER-ALPHA-TRYPESIN A	2.15e+02
40	19	100.0	928	1	ATCI_YARLI CALCIUM-TRANSPORTIN IN	2.15e+02
41	19	100.0	996	1	ATNA_ARTSA SODIUM/POTASSIUM-TRANS	2.15e+02
42	19	100.0	1100	1	JAK3_RAT TYROSINE-PROTEIN KINAS	2.15e+02
43	19	100.0	1177	1	JAK_DROME TYROSINE-PROTEIN KINAS	2.15e+02
44	19	100.0	1490	1	CDR4_CANAL ABC TRANSPORTER CDR4	2.15e+02
45	19	100.0	2749	1	IP3R_RAT INOSITOL 1,4,5-TRISPHO	2.15e+02

## ALIGNMENTS

RESULT 1  
ID CHLB\_OSMCL STANDARD; PRT; 103 AA.  
AC P37851;  
DT 01-OCT-1994 (Rel. 30, Created)  
DT 01-OCT-1994 (Rel. 30, Last sequence update)  
DT 01-OCT-1994 (Rel. 30, Last annotation update)  
DE PROTOCHLOROPHYLLIDE REDUCTASE CHLB SUBUNIT (EC 1.3.1.33) (NADPH-  
DE PROTOCHLOROPHYLLIDE OXIDOREDUCTASE CHLB SUBUNIT) (FRAGMENT).  
GN CHLB.  
OS Osmunda claytoniana (Fern).

OC Chloroplast.  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC euphyllophytes; Filicophyta; Filicopsida; Filicales; Osmundaceae;  
OC Osmunda.  
RN [1]

RP SEQUENCE FROM N.A.

RA BOIVIN R., RICHARD M., BOUSQUET J., BELLEMAIRE G.;

RL Submitted (XXX-1994) to the EMBL/GenBank/DBJ databases

CC -!- FUNCTION: INVOLVED IN THE LIGHT-INDEPENDENT ACCUMULATION OF

CC CHLOROPHYLL, PROBABLY AT THE STEP OF REDUCTION OF PROTO-

CC CHLOROPHYLLIDE TO CHLOROPHYLLIDE (BY SIMILARITY).

CC -!- CATALYTIC ACTIVITY: CHLOROPHYLLIDE A + NADP(+) -

CC PROTOCHLOROPHYLLIDE + NADPH.

CC -!- PATHWAY: LIGHT-INDEPENDENT CHLOROPHYLL BIOSYNTHESIS.

CC -!- SIMILARITY: BELONGS TO THE CHLB / BCHB FAMILY.

CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).

CC EMBL: L25771; AAC37492.1; -

CC MENDEL; 2282; OSMCL:chlB.1.

CC PFAM: PF00148; Oxidored\_nitro; 1.

CC Chloroplast; Photosynthesis; Chlorophyll biosynthesis; Oxidoreductase;

CC NADP.

FT NON\_TER 103

SQ SEQUENCE 103 AA; 11661 MW; A43CD11E CRC32;

Query Match 100.0%; Score 19; DB 1; Length 103;  
Best Local Similarity 25.0%; Pred. No. 2.15e+02;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;



```

GN AF1809.
OS Archaeoglobus fulgidus.
OC Archaea, Euryarchaeota; Archaeoglobales; Archaeoglobaceae;
OC Archaeoglobus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE; 98049343.
RA KLENK H.-P., CLAYTON R.A., TOMB J.-F., WHITE O., NELSON K.E.,
RA KETCHUM K.A., DODSON R.J., GWINN M., HICKEY E.K., PETERSON J.D.,
RA RICHARDSON D.L., KERAVAGE A.R., GRAHAM D.E., KIRPIDES N.C.,
RA FLEISCHMANN R.D., OURAKENBUSH J., LEE N.H., SUTTON G.G., GILL S.,
RA KIRKNESS E.F., DOUGHERTY B.A., MCKENNEY K., ADAMS M.D., LOFTUS B.,
RA PETERSON S., REICH C.I., MCNEIL L.K., BADGER J.H., GLODEK A., ZHOU L.,
RA OVERBEER R., GOCAYNE J.D., WEIDMAN J.F., McDONALD L., UTTERBACK T.,
RA COTTON M.D., SPRIGGS T., ARTIACH P., KAINE B.P., SYKES S.M.,
RA SADOW P.W., D'ANDREA K.P., BOWMAN C., FUJII C., GARLAND S.A.,
RA MASON T.M., OLSEN G.J., FRASER C.M., SMITH H.O., WOESE C.R.,
RA VENTER J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
RT reducing archaeon Archaeoglobus fulgidus.";
RL Nature 390:364-370(1997).
DR EMBL; AE000978; AAB89452.1; -.
DR TIGR; AF1809; -.
KW Hypothetical protein.
SQ SEQUENCE 88 AA; 9875 MW; A3D560FF CRC32;

Query Match 100.0%; Score 16; DB 1; Length 88;
Best Local Similarity 25.0%; Pred. No. 7.96e+02;
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 62 PVTLFRTL 69
QY 2 PXXXXXXL 9

RESULT 3
ID P94628 PRELIMINARY; PRT; 95 AA.
AC P94628;
DT 01-MAY-1997 (TREMBLrel. 03, Created)
DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)
DE 01-MAY-1997 (TREMBLrel. 03, Last annotation update)
DE DNA INTEGRASE (FRAGMENT).
GN INT.
OS Citrobacter freundii.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Citrobacter.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K1.
RX MEDLINE; 99310339.
RA KAWARABAYASI Y., HINO Y., HORIKAWA H., YAMAZAKI S., HAIKAWA Y.,
RA JIN-NO K., TAKAHASHI M., SEKINE M., BABA S., ANKAI A., KOSUGI H.,
RA HOSOYAMA A., FUKUI S., NAGAI Y., NISHIJIMA K., NAKAZAWA H.,
RA TAKAMIYA M., MASUDA S., FUNAHASHI T., TANAKA T., KUDOH Y.,
RA YAMAZAKI J., KUSHIDA N., OGUCHI A., AOKI K., KUBOTA K., NAKAMURA Y.,
RA NOMURA N., SAKO Y., KIKUCHI H.;
RT "Complete genome sequence of an aerobic hyper-thermophilic
RT crenarchaeon, Aeropyrum pernix K1.";
RL DNA Res. 6:83-101(1999).
DR EMBL; AP000060; BAA79675.1; -.
DR EMBL; AP000060; BAA79675.1; -.
SQ SEQUENCE 101 AA; 10251 MW; 577A34B7 CRC32;

Query Match 100.0%; Score 16; DB 2; Length 95;
Best Local Similarity 25.0%; Pred. No. 7.96e+02;
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 10 PLRSVKVL 17
QY 2 PXXXXXXL 9

RESULT 4
ID Q9YAX3 PRELIMINARY; PRT; 100 AA.
AC Q9YAX3;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
DE 100AA LONG HYPOTHETICAL PROTEIN.
GN APE1822.

OS Aeropyrum pernix.
OC Archaea; Crenarchaeota; Aeropyrum.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K1.
RX MEDLINE; 99310339.
RA KAWARABAYASI Y., HINO Y., HORIKAWA H., YAMAZAKI S., HAIKAWA Y.,
RA JIN-NO K., TAKAHASHI M., SEKINE M., BABA S., ANKAI A., KOSUGI H.,
RA HOSOYAMA A., FUKUI S., NAGAI Y., NISHIJIMA K., NAKAZAWA H.,
RA TAKAMIYA M., MASUDA S., FUNAHASHI T., TANAKA T., KUDOH Y.,
RA YAMAZAKI J., KUSHIDA N., OGUCHI A., AOKI K., KUBOTA K., NAKAMURA Y.,
RA NOMURA N., SAKO Y., KIKUCHI H.;
RT "Complete genome sequence of an aerobic hyper-thermophilic
RT crenarchaeon, Aeropyrum pernix K1.";
RL DNA Res. 6:83-101(1999).
DR EMBL; AP000062; BAA80825.1; -.
DR EMBL; AP000062; BAA80825.1; -.
SQ SEQUENCE 100 AA; 11180 MW; 911DBEE3 CRC32;

Query Match 100.0%; Score 16; DB 1; Length 100;
Best Local Similarity 25.0%; Pred. No. 7.96e+02;
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 28 PGLVEPL 35
QY 2 PXXXXXXL 9

RESULT 5
ID Q9YE71 PRELIMINARY; PRT; 101 AA.
AC Q9YE71;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
DE 101AA LONG HYPOTHETICAL PROTEIN.
GN APE0699.
OS Aeropyrum pernix.
OC Archaea; Crenarchaeota; Aeropyrum.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K1.
RX MEDLINE; 99310339.
RA KAWARABAYASI Y., HINO Y., HORIKAWA H., YAMAZAKI S., HAIKAWA Y.,
RA JIN-NO K., TAKAHASHI M., SEKINE M., BABA S., ANKAI A., KOSUGI H.,
RA HOSOYAMA A., FUKUI S., NAGAI Y., NISHIJIMA K., NAKAZAWA H.,
RA TAKAMIYA M., MASUDA S., FUNAHASHI T., TANAKA T., KUDOH Y.,
RA YAMAZAKI J., KUSHIDA N., OGUCHI A., AOKI K., KUBOTA K., NAKAMURA Y.,
RA NOMURA N., SAKO Y., KIKUCHI H.;
RT "Complete genome sequence of an aerobic hyper-thermophilic
RT crenarchaeon, Aeropyrum pernix K1.";
RL DNA Res. 6:83-101(1999).
DR EMBL; AP000060; BAA79675.1; -.
DR EMBL; AP000060; BAA79675.1; -.
SQ SEQUENCE 101 AA; 10794 MW; F221BD95 CRC32;

Query Match 100.0%; Score 16; DB 1; Length 101;
Best Local Similarity 25.0%; Pred. No. 7.96e+02;
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 5 PSLWTFLL 12
QY 2 PXXXXXXL 9

RESULT 6
ID O29589 PRELIMINARY; PRT; 101 AA.
AC O29589;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-AUG-1998 (TREMBLrel. 07, Last annotation update)
DE CONSERVED HYPOTHETICAL PROTEIN.
GN AF0668.
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobales; Archaeoglobaceae;
OC Archaeoglobus.

```

```

RN  SEQUENCE FROM N.A.
RP  STRAIN-VC-16 / DSM 4304 / ATCC 49558;
RX  MEDLINE; 98049343
RA  KLENK H.-P., CLAYTON R.A., TOMB J.-F., WHITE O., NELSON K.E.,
RA  KETCHUM K.A., DODSON R.J., GWINN M., HICKEY E.K., PETERSON J.D.,
RA  RICHARDSON D.L., KERLAVAGE A.R., GRAHAM D.E., KYRPIDES N.C.,
RA  FEISCHMANN R.D., QUACKENBUSH J., LEE N.H., SUTTON G.G., GILL S.,
RA  KIRKNESS E.F., DOUGHERTY B.A., MCKENNEY K., ADAMS M.D., LOFTUS B.,
RA  PETERSON S., REICH C.I., MCNEIL L.K., BADGER J.H., GLODEK A., ZHOU L.,
RA  OVERBECK R., GOCAYNE J.D., WEIDMAN J.F., McDONALD L., UTTERBACK T.,
RA  COTTON M.D., SPRIGGS T., ARTIACH P., KAINE B.P., SYKES S.M.,
RA  SADOW P.W., D'ANDREA K.P., BOWMAN C., FUJII C., GARLAND S.A.,
RA  MASON T.N., OLSEN G.J., FRASER C.M., SMITH H.O., WOESE C.R.,
RA  VENTER J.C.;
RT  "The complete genome sequence of the hyperthermophilic, sulphate-
RT  reducing archaeon Archaeoglobus fulgidus.";
RL  Nature 390:364-370(1997).
DR  EMBL; AE001058; AAB90572.1; -.
KW  TIGR; AF0668; -.
SQ  SEQUENCE 101 AA; 11230 MW; 9EC9BD2E CRC32;

Query Match 100.0%; Score 16; DB 1; Length 101;
Best Local Similarity 25.0%; Pred. No. 7.96e+02;
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 92 PDDAVEDL 99
QY 2 PXXXXXXL 9

RESULT 7
ID Q9Y46 PRELIMINARY; PRT; 108 AA.
AC Q9Y46
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE 108AA LONG HYPOTHETICAL PROTEIN.
GN APE0724.
OS Aeropyrum pernix.
OC Archaea; Crenarchaeota; Aeropyrum.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-K1;
RX MEDLINE; 99310339.
RA KAWARABAYASI Y., HINO Y., HORIKAWA H., YAMAZAKI S., HAIKAWA Y.,
RA JIN-NO K., TAKAHASHI M., SEKINE M., NAGAI Y., NISHIJIMA K., KOSUGI H.,
RA HOSOVAYA A., FUKUI S., NAGAI Y., NISHIJIMA K., NAKAZAWA H.,
RA TAKAMIYA M., MASUDA S., FUNAHASHI T., TANAKA T., KUDOH Y.,
RA YAMAZAKI J., KUSHIDA N., OGUCHI A., AOKI K., KUBOTA K., NAKAMURA Y.,
RA NOMURA N., SAKO Y., KIKUCHI H.;
RT "Complete genome sequence of an aerobic hyper-thermophilic
RT crenarchaeon, Aeropyrum pernix K1.";
RL DNA Res. 6:83-101(1999).
DR EMBL; AP000060; BAA79700.1; -.
SQ SEQUENCE 108 AA; 11190 MW; BDA66C20 CRC32;

Query Match 100.0%; Score 16; DB 1; Length 108;
Best Local Similarity 25.0%; Pred. No. 7.96e+02;
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 62 PSVRVGL 69
QY 2 PXXXXXXL 9

RESULT 8
ID Q59492 PRELIMINARY; PRT; 115 AA.
AC Q59492;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-JAN-1999 (TrEMBLrel. 09, Last annotation update)

```

```

DE 115AA LONG HYPOTHETICAL PROTEIN.
GN PH1828.
OS Pyrococcus horikoshii.
OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-OT3;
RX MEDLINE; 98344137.
RA KAWARABAYASI Y., SAWADA M., HORIKAWA H., HAIKAWA Y., HINO Y.,
RA YAMAMOTO S., SEKINE M., BABA S., KOSUGI H., HOSOVAYA A., NAGAI Y.,
RA SAKAI M., OGURA K., OTUKA R., NAKAZAWA H., TAKAMIYA M., OHFUKU Y.,
RA FUNAHASHI T., TANAKA T., KUDOH Y., YAMAZAKI J., KUSHIDA N., OGUCHI A.,
RA AOKI K., NAKAMURA Y., ROBB T.F., HORIKOSHI K., MASUCHI Y., SHIZUYA H.,
RA KIKUCHI H.;
RT "Complete sequence and gene organization of the genome of a hyper-
RT thermophilic archaeobacterium, Pyrococcus horikoshii OT3.";
RL DNA Res. 5:55-76(1998).
DR EMBL; AP000007; BAA30947.1; -.
SQ SEQUENCE 115 AA; 12209 MW; 1846C3BA CRC32;

Query Match 100.0%; Score 16; DB 1; Length 115;
Best Local Similarity 25.0%; Pred. No. 7.96e+02;
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 51 PIEIVRTL 58
QY 2 PXXXXXXL 9

RESULT 9
ID Q9YAX4 PRELIMINARY; PRT; 129 AA.
AC Q9YAX4;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE 129AA LONG HYPOTHETICAL PROTEIN.
GN APE1821.
OS Aeropyrum pernix.
OC Archaea; Crenarchaeota; Aeropyrum.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-K1;
RX MEDLINE; 99310339.
RA KAWARABAYASI Y., HINO Y., HORIKAWA H., YAMAZAKI S., HAIKAWA Y.,
RA JIN-NO K., TAKAHASHI M., SEKINE M., BABA S., ANKAI A., KOSUGI H.,
RA HOSOVAYA A., FUKUI S., NAGAI Y., NISHIJIMA K., NAKAZAWA H.,
RA TAKAMIYA M., MASUDA S., FUNAHASHI T., TANAKA T., KUDOH Y.,
RA YAMAZAKI J., KUSHIDA N., OGUCHI A., AOKI K., KUBOTA K., NAKAMURA Y.,
RA NOMURA N., SAKO Y., KIKUCHI H.;
RT "Complete genome sequence of an aerobic hyper-thermophilic
RT crenarchaeon, Aeropyrum pernix K1.";
RL DNA Res. 6:83-101(1999).
DR EMBL; AP000062; BAA80824.1; -.
SQ SEQUENCE 129 AA; 14887 MW; A3718E42 CRC32;

Query Match 100.0%; Score 16; DB 1; Length 129;
Best Local Similarity 25.0%; Pred. No. 7.96e+02;
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 55 PVTFPSHL 62
QY 2 PXXXXXXL 9

RESULT 10
ID Q9YB06 PRELIMINARY; PRT; 131 AA.
AC Q9YB06;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE 131AA LONG HYPOTHETICAL PROTEIN.
GN APE1789.
OS Aeropyrum pernix.

```

```
OC Archaea; Crenarchaeota; Aeropyrum.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K1;
RX MEDLINE; 99310339.
RA KAWARABAYASI Y., HINO Y., HORIKAWA H., YAMAZAKI S., HAIKAWA Y.,
RA JIN-NO K., TAKAHASHI M., SEKINE M., BABA S., ANKAI A., KOSUGI H.,
RA HOSOVAMA A., FUKUI S., NAGAI Y., NISHIJIMA K., NAKAZAWA H.,
RA TAKAMIYA M., MASUDA S., FUNAHASHI T., TANAKA T., KUDOH Y.,
RA YAMAZAKI J., KUSHIDA N., OGUCHI A., AOKI K., KUBOTA K., NAKAMURA Y.,
RA NOMURA N., SAKO Y., KIKUCHI H.;
RT "Complete genome sequence of an aerobic hyper-thermophilic
RT crenarchaeon, Aeropyrum pernix K1.";
RL DNA Res. 6:83-101(1999).
DR EMBL; AP000062; BAA80792.1; -
SQ SEQUENCE 131 AA; 14568 MW; 6B1ABAD8 CRC32;

Query Match 100.0%; Score 16; DB 1; Length 131;
Best Local Similarity 25.0%; Pred. No. 7.96e+02;
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 50 PPVPSKAL 57
QY 2 PXXXXXXL 9

RESULT 11
ID O26321 PRELIMINARY; PRT; 137 AA.
AC O26321;
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DT 01-AUG-1998 (TREMBlrel. 07, Last annotation update)
DE HYPOTHETICAL 15.9 KD PROTEIN.
GN MTH219.
OS Methanobacterium thermoautotrophicum.
OC Archaea; Euryarchaeota; Methanobacteriales; Methanobacteriaceae;
OC Methanobacterium.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-DELTA H;
RX MEDLINE; 96037514.
RA SMITH D.R., DOUCETTE-STAMM L.A., DELOUGHERY C., LEE H.-M., DUBOIS J.,
RA ALDREDGE T., BASHIRZADEH R., BLAKELY D., COOK R., GILBERT K.,
RA HARRISON D., HOANG L., KEAGLE P., LUW W., POTHIER B., QIU D.,
RA SPADAFORA R., VICARE R., WANG Y., WIERZBOWSKI J., GIBSON R.,
RA JIWANI N., CARUSO A., BUSH D., SAPER H., PATWELL D., PRASHAKAR S.,
RA MCDUGALL S., SHIMER G., GOYAL A., PIETROVSKI S., CHURCH G.M.,
RA DANIELS C.J., MAO J.-I., RICE P., NOLLING J., REEVE J.N.;
RT "Complete genome sequence of Methanobacterium thermoautotrophicum
RT deltaH: functional analysis and comparative genomics.";
RL J. Bacteriol. 179:7135-7155(1997).
DR EMBL; AF000809; AAB84725.1; -
KW Hypothetical protein.
SQ SEQUENCE 137 AA; 15932 MW; BB25D06B CRC32;

Query Match 100.0%; Score 16; DB 1; Length 137;
Best Local Similarity 25.0%; Pred. No. 7.96e+02;
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 123 PDFCOTFL 130
QY 2 PXXXXXXL 9

RESULT 12
ID Q9YBS3 PRELIMINARY; PRT; 141 AA.
AC Q9YBS3;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
DE 141AA LONG HYPOTHETICAL
DE CDP-DIACYLGLYCEROL--GLYCEROL-3--PHOSPHATIDYLTTRANSFERASE.
GN APE1526.

OC Archaea; Crenarchaeota; Aeropyrum.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K1;
RX MEDLINE; 99310339.
RA KAWARABAYASI Y., HINO Y., HORIKAWA H., YAMAZAKI S., HAIKAWA Y.,
RA JIN-NO K., TAKAHASHI M., SEKINE M., BABA S., ANKAI A., KOSUGI H.,
RA HOSOVAMA A., FUKUI S., NAGAI Y., NISHIJIMA K., NAKAZAWA H.,
RA TAKAMIYA M., MASUDA S., FUNAHASHI T., TANAKA T., KUDOH Y.,
RA YAMAZAKI J., KUSHIDA N., OGUCHI A., AOKI K., KUBOTA K., NAKAMURA Y.,
RA NOMURA N., SAKO Y., KIKUCHI H.;
RT "Complete genome sequence of an aerobic hyper-thermophilic
RT crenarchaeon, Aeropyrum pernix K1.";
RL DNA Res. 6:83-101(1999).
DR EMBL; AP000062; BAA80792.1; -
SQ SEQUENCE 131 AA; 14568 MW; 6B1ABAD8 CRC32;

Query Match 100.0%; Score 16; DB 1; Length 131;
Best Local Similarity 25.0%; Pred. No. 7.96e+02;
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 50 PPVPSKAL 57
QY 2 PXXXXXXL 9

RESULT 11
ID O26321 PRELIMINARY; PRT; 137 AA.
AC O26321;
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DT 01-AUG-1998 (TREMBlrel. 07, Last annotation update)
DE HYPOTHETICAL 15.9 KD PROTEIN.
GN MTH219.
OS Methanobacterium thermoautotrophicum.
OC Archaea; Euryarchaeota; Methanobacteriales; Methanobacteriaceae;
OC Methanobacterium.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-DELTA H;
RX MEDLINE; 96037514.
RA SMITH D.R., DOUCETTE-STAMM L.A., DELOUGHERY C., LEE H.-M., DUBOIS J.,
RA ALDREDGE T., BASHIRZADEH R., BLAKELY D., COOK R., GILBERT K.,
RA HARRISON D., HOANG L., KEAGLE P., LUW W., POTHIER B., QIU D.,
RA SPADAFORA R., VICARE R., WANG Y., WIERZBOWSKI J., GIBSON R.,
RA JIWANI N., CARUSO A., BUSH D., SAPER H., PATWELL D., PRASHAKAR S.,
RA MCDUGALL S., SHIMER G., GOYAL A., PIETROVSKI S., CHURCH G.M.,
RA DANIELS C.J., MAO J.-I., RICE P., NOLLING J., REEVE J.N.;
RT "Complete genome sequence of Methanobacterium thermoautotrophicum
RT deltaH: functional analysis and comparative genomics.";
RL J. Bacteriol. 179:7135-7155(1997).
DR EMBL; AF000809; AAB84725.1; -
KW Hypothetical protein.
SQ SEQUENCE 137 AA; 15932 MW; BB25D06B CRC32;

Query Match 100.0%; Score 16; DB 1; Length 137;
Best Local Similarity 25.0%; Pred. No. 7.96e+02;
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 123 PDFCOTFL 130
QY 2 PXXXXXXL 9

RESULT 12
ID Q9YBS3 PRELIMINARY; PRT; 141 AA.
AC Q9YBS3;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
DE 141AA LONG HYPOTHETICAL
DE CDP-DIACYLGLYCEROL--GLYCEROL-3--PHOSPHATIDYLTTRANSFERASE.
GN APE1526.

OS Aeropyrum pernix.
OC Archaea; Crenarchaeota; Aeropyrum.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K1;
RX MEDLINE; 99310339.
RA KAWARABAYASI Y., HINO Y., HORIKAWA H., YAMAZAKI S., HAIKAWA Y.,
RA JIN-NO K., TAKAHASHI M., SEKINE M., BABA S., ANKAI A., KOSUGI H.,
RA HOSOVAMA A., FUKUI S., NAGAI Y., NISHIJIMA K., NAKAZAWA H.,
RA TAKAMIYA M., MASUDA S., FUNAHASHI T., TANAKA T., KUDOH Y.,
RA YAMAZAKI J., KUSHIDA N., OGUCHI A., AOKI K., KUBOTA K., NAKAMURA Y.,
RA NOMURA N., SAKO Y., KIKUCHI H.;
RT "Complete genome sequence of an aerobic hyper-thermophilic
RT crenarchaeon, Aeropyrum pernix K1.";
RL DNA Res. 6:83-101(1999).
DR EMBL; AP000061; BAA80525.1; -
DR PROSITE; PS00379; CDP-ALCOHOL_P_TRANSF; 1.
KW Transferase.
SQ SEQUENCE 141 AA; 15031 MW; 5E9191C4 CRC32;

Query Match 100.0%; Score 16; DB 1; Length 141;
Best Local Similarity 25.0%; Pred. No. 7.96e+02;
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 57 PLIVIAML 64
QY 2 PXXXXXXL 9

RESULT 13
ID O54612 PRELIMINARY; PRT; 150 AA.
AC O54612;
DT 01-JUN-1998 (TREMBlrel. 06, Created)
DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)
DT 01-JUN-1998 (TREMBlrel. 06, Last annotation update)
DE PHOSPHOGLUCOMUTASE (FRAGMENT).
GN PGN.
OS Neisseria meningitidis.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
RN [1]
RP SEQUENCE FROM N.A.
RA MAIDEN M.C.J., BYGRAVES J.A., FEIL E., MORELLI G., RUSSELL J.E.,
RA URWIN R., ZHANG Q., ZHOU J., ZURTH K., CAUGANT D.A., FEAVERS I.M.,
RA ACHTMAN M., SPRATT B.G.;
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF037871; AAC08874.1; -
DR EMBL; AF037862; AAC08865.1; -
DR EMBL; AF037869; AAC08872.1; -
FT NON_TER 1
FT NON_TER 150
SQ SEQUENCE 150 AA; 16510 MW; A5A2B4F4 CRC32;

Query Match 100.0%; Score 16; DB 2; Length 150;
Best Local Similarity 25.0%; Pred. No. 7.96e+02;
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 110 PSEVLNLL 117
QY 2 PXXXXXXL 9

RESULT 14
ID O30747 PRELIMINARY; PRT; 164 AA.
AC O30747;
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DT 01-JAN-1998 (TREMBlrel. 05, Last annotation update)
DE NNRS (FRAGMENT).
GN NNRS.
OS Rhodobacter sphaeroides (Rhodopseudomonas sphaeroides).
OC Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;
OC Rhodobacter.
RN [1]
```



RP SEQUENCE FROM N.A.  
RC STRAIN-2.4.1;  
RA KWATKOWSKI A.V., LARATTA W.P., TOFFANIN A., SHAPLEIGH J.P.;  
RL J. Bacteriol. 0:0-0(1997).  
DR EMBL: AF016258; AAB69131.1; -  
FT NON\_TER 164 164  
SQ SEQUENCE 164 AA; 17552 MW; 35D357E5 CRC32;

Query Match 100.0%; Score 16; DB 2; Length 164;  
Best Local Similarity 25.0%; Pred. No. 7.96e+02;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 21 PFLLSAL 28  
|  
Qy 2 PXXXXXXL 9

RESULT 15  
ID 007858 PRELIMINARY; PRT: 167 AA.  
AC 007858;  
DT 01-JUL-1997 (TRENBLrel. 04, Created)  
DT 01-JUL-1997 (TRENBLrel. 04, Last sequence update)  
DT 01-AUG-1998 (TRENBLrel. 07, Last annotation update)  
DE TRANSPOSASE.  
GN INSB.  
OS Shigella boydii.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Shigella.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA CHEN J.H., HSIU W.B.;  
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL: U96388; AAB61273.1; -  
SQ SEQUENCE 167 AA; 19737 MW; EFDF0EEB CRC32;

Query Match 100.0%; Score 16; DB 2; Length 167;  
Best Local Similarity 25.0%; Pred. No. 7.96e+02;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 12 PQHGFTSL 19  
|  
Qy 2 PXXXXXXL 9

Search completed: Sat Apr 15 02:13:44 2000  
Job time : 90 secs.

THIS PAGE BLANK (USPTO)

\*\*\*\*\*  
M P S R L  
\*\*\*\*\*  
(TM)

\*\*\*\*\*  
Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd  
\*\*\*\*\*  
Mpsrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Sat Apr 15 02:16:52 2000; Maspar time 3.18 Seconds  
Tabular output not generated. 66.966 Million cell updates/sec

Title: >US-08-452-843-30  
Description: (1-9) from US08452843.pep  
Sequence: 1 XPXXXXXX 9  
Scoring table: PAM 150  
Gap 15  
Searched: 188963 seqs, 23686106 residues  
Post-processing: Minimum Match 0%  
Listing first 45 summaries  
Database: a-geneseq36  
1:geneseqp  
Statistics: Mean 8.320; Variance 13.909; scale 0.598

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES				Pred. No.	
Result No.	Score	Query Match	Description	ID	
1	19	100.0	Substance P [Me Gly 6,	R21962	1.96e+03
2	19	100.0	Calmodulin inhibiting	R39552	1.96e+03
3	19	100.0	HLA binding peptide ho	R33691	1.96e+03
4	19	100.0	Pro-LAP.	R86895	1.96e+03
5	19	100.0	Mutant serine protease	R88973	1.96e+03
6	19	100.0	Mutant serine protease	R88976	1.96e+03
7	19	100.0	CRL SCR-2.	R72333	1.96e+03
8	19	100.0	Variant IGE - mutant E	R32036	1.96e+03
9	19	100.0	Variant IGE - mutant E	R32035	1.96e+03
10	19	100.0	Lymphotoxin protein.	R22037	1.96e+03
11	19	100.0	Plasminostreptin gene	R40945	1.96e+03
12	19	100.0	MHS2:MnSOD variant.	R60361	1.96e+03
13	19	100.0	Interleukin 4 componen	R82934	1.96e+03
14	19	100.0	Plasmod PRTA735-encode	R11178	1.96e+03
15	19	100.0	M. tuberculosis inhA.	R65901	1.96e+03
16	19	100.0	Subtilisin 309 Serine	R46428	1.96e+03
17	19	100.0	Subtilisin 309 Serine	R46426	1.96e+03
18	19	100.0	Subtilisin 309 Serine	R87997	1.96e+03
19	19	100.0	Subtilisin 309 loop 4	R28382	1.96e+03
20	19	100.0	Mutant subtilisin 309	R25538	1.96e+03
21	19	100.0	Subtilisin protease.	R46360	1.96e+03
22	19	100.0	Subtilisin 309 Serine	R46361	1.96e+03
23	19	100.0	PB92 serine protease,	R10985	1.96e+03
			Dehalogenase.		

24	19	100.0	365	1	R44235	HTLE receptor.	1.96e+03
25	19	100.0	436	1	R14487	Soluble interferon- $\alpha$ p	1.96e+03
26	19	100.0	474	1	R59866	Human GABA receptor be	1.96e+03
27	19	100.0	483	1	R57993	M138X alpha-amylase.	1.96e+03
28	19	100.0	505	1	R40859	Flavonoid-3',5'-hydrox	1.96e+03
29	19	100.0	510	1	R41195	Yeast delta 9 desatura	1.96e+03
30	19	100.0	511	1	R40842	SHPP.	1.96e+03
31	19	100.0	520	1	R85943	Human-1B adrenergic re	1.96e+03
32	19	100.0	528	1	R63438	Human enteric alkaline	1.96e+03
33	19	100.0	575	1	R22395	AOAH encoded by th	1.96e+03
34	19	100.0	690	1	R40306	Sequence encoded by th	1.96e+03
35	19	100.0	720	1	R15381	Pseudomonas S177-glu	1.96e+03
36	19	100.0	738	1	R69849	Ethylene response (ETR	1.96e+03
37	19	100.0	770	1	R13786	HIV multifunctional fu	1.96e+03
38	19	100.0	849	1	R63070	Human EAA3d excitatory	1.96e+03
39	19	100.0	860	1	R92716	Mouse muscle-localised	1.96e+03
40	19	100.0	1023	1	R15054	Polypeptide encoded GL	1.96e+03
41	19	100.0	1367	1	R67537	Mouse flk-1.	1.96e+03
42	19	100.0	1433	1	R39568	Sequence of c-erbB-2 t	1.96e+03
43	19	100.0	1480	1	R13302	CFTR L1077P.	1.96e+03
44	19	100.0	1684	1	R14948	Bacterial amylase A-18	1.96e+03
45	19	100.0	3054	1	R40841	Translation of TEV lar	1.96e+03

ALIGNMENTS

RESULT 1  
ID R21962 standard; Peptide; 11 AA.  
AC R21962;  
DT 25-JUN-1992 (first entry)  
DE Substance P [Me Gly 6, Met (O2) 11].  
KW Tachykinin agonist; beta-amyloid; Alzheimer's disease; Down's;  
KW syndrome; hereditary cerebral haemorrhage.  
OS Synthetic.  
PH Key Location/Qualifiers  
FT misc\_difference 6  
FT /label= OTHER  
FT /note= "OTHER - Methyl glycine"  
FT misc\_difference 11  
FT /label= OTHER  
FT /note= "OTHER - Met (O2)"  
PN W09202248-A.  
PD 20-FEB-1992.  
PF 29-JUL-1991; U05323.  
PR 27-JUL-1990; US-559173.  
PA (CHIL-) CHILDRENS MED CENT.  
PI Yankner BA;  
DR WPI; 92-079804/10.  
PT Treatment of neuronal accumulation of beta-amyloid - using  
PT tachykinin agonists e.g. substance P, physalaemin and neurokinin  
PT B. for treating Alzheimer's disease, Downs syndrome, etc.  
PS Claim 10; Page 22; 35pp; English.  
CC The peptide is the tachykinin agonist, substance P with methyl  
CC glycine substituted at position 9 and Met (O2) at position 11.  
CC The peptide was synthesised by standard solid phase synthesis.  
CC Neuronal accumulation of beta-amyloid may be treated by administ-  
CC ration of tachykinin agonists. The peptide can reduce the neuro-  
CC toxic effects of a beta-amyloid related polypeptide on cultured  
CC neurons. The peptide and its analogues are useful for controlling  
CC diseases characterised by beta amyloid accumulation in the brain  
CC such as Alzheimer's disease and Down's syndrome.  
CC See also R21932-75.  
SQ Sequence 11 AA;

Query Match 100.0%; Score 19; DB 1; Length 11;  
Best Local Similarity 25.0%; Pred. No. 1.96e+03;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 4 PQOFFGLM 11  
Qy 2 PXXXXXX 9

RESULT 2  
ID R39552 standard; peptide; 15 AA.  
AC R39552;  
DT 10-FEB-1994 (first entry)  
DE Calmodulin inhibiting peptide #29.  
KW Calmodulin-binding peptide; alpha-helix; cell growth; infection;  
KW cell motility; pregnancy; carbohydrate metabolism; tumours;  
KW neurotransmission; anti-psychotic activity; anti-inflammatory activity;  
KW hyperproliferation.  
OS Synthetic.  
PN W09316100-A.  
PD 19-AUG-1993.  
PF 08-FEB-1993; U01112.  
PR 06-FEB-1992; US-831219.  
PA (UYCI-) UNIV CINICINNAZI.  
PI Dedman JR, Jamieson GA, Keetzel MA;  
DR WPI; 93-272825/34.  
PT Calmodulin-binding peptide(s) and derivs. - useful for inhibiting  
PT calmodulin, esp. in treatment of hyper proliferative diseases  
PT e.g. cancer  
PS Claim 1; Page 13; 18pp; English.  
CC The sequences given in R39552-52 are calmodulin-binding peptides.  
CC These peptides are unusual as they do not adopt a strict alpha-helix,  
CC as do most calmodulin-binding peptides and many of them also  
CC contain a Trp residue juxtaposed to a pro residue. These peptides  
CC are useful for treating tumours or infection, for control of cell  
CC growth, division and meiosis, regulation of cell motility,  
CC prevention of pregnancy, regulation of carbohydrate metabolism,  
CC regulation of neurotransmission, anti-psychotic activity,  
CC anti-inflammatory activity and especially hyperproliferative  
CC disorders.  
SQ Sequence 15 AA;

Query Match 100.0%; Score 19; DB 1; Length 15;  
Best Local Similarity 25.0%; Pred. No. 1.96e+03;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 5 PNLTLQRM 12  
| |  
QY 2 PXXXXXXM 9

RESULT 3  
ID R83691 standard; peptide; 20 AA.  
AC R83691;  
DT 10-APR-1996 (first entry)  
DE HLA binding peptide homologous to CD20 (B lymphocyte antigen).  
KW HLA binding oligopeptide; immunosuppressant; autoimmune disease;  
KW CD20; B lymphocyte antigen; residues 26-45; homologue.  
OS Synthetic.  
PN J07206896-A.  
PD 08-AUG-1995.  
PF 20-JAN-1994; 004615.  
PR 20-JAN-1994; JP-004615.  
PA (TEIJ ) TEIJIN LTD.  
DR WPI; 95-309097/40.  
PT New HLA binding oligo-peptide(s) - useful as immunosuppressants for  
PT treating auto-immune diseases  
PS Example 1; Page 5; 9pp; Japanese.  
CC The present peptide is homologous to the CD20 (B lymphocyte antigen)  
CC residues 26-45, and is a HLA binding oligopeptide. It can be used as  
CC an immunosuppressant for the treatment of autoimmune diseases.  
SQ Sequence 20 AA;

Query Match 100.0%; Score 19; DB 1; Length 20;  
Best Local Similarity 25.0%; Pred. No. 1.96e+03;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 2 PKPLFRRM 9  
| |  
QY 2 PXXXXXXM 9

RESULT 4  
ID R86895 standard; peptide; 44 AA.  
AC R86895;  
DT 26-JUN-1996 (first entry)  
DE Pro-LAP.  
KW Lingual antimicrobial peptide; epithelium; LAP; bovine; beta-defensin;  
KW antimicrobial activity; Gram-positive bacteria; Gram negative bacteria;  
KW fungal pathogen; mammal; microbial infection; immunodeficiency; AIDS;  
KW cystic fibrosis; gum disease; burn; pneumonia.  
OS Bos taurus.  
FH Key Location/Qualifiers  
FT peptide 1..2 /note= "pro region"  
FT peptide 3..44 /note= "mature LAP"  
PN W09532287-A1.  
PD 30-NOV-1995.  
PF 24-MAY-1995; U06761.  
PR 24-MAY-1994; US-248016.  
PA (MAGA-) MAGAININ PHARM INC.  
PI Schonvetter BS, Zasloff MA;  
DR WPI; 96-020582/02.  
DR N-P8DB; T07134.  
PT New antimicrobial peptide from mammalian lingual epithelium - for  
PT treating bacterial or fungal infections, esp. of the epithelium,  
PT e.g. in gum disease, cystic fibrosis, burns, etc.  
PS Claim 8; Page 21; 37pp; English.  
CC This sequence represents the pro form of bovine epithelial lingual  
CC antimicrobial peptide (LAP). LAP is a member of the beta-defensin group  
CC of the defensin family of peptides. LAP has broad spectrum antimicrobial  
CC activity against Gram-positive and Gram negative bacteria, and fungal  
CC pathogens. LAP is present at low levels in mammalian epithelia, with  
CC high expression levels being induced in response to epithelial  
CC injury/infection. The cDNA encoding this sequence can be used in a  
CC method of identifying endogenous up-regulators of LAP. In this method,  
CC epithelial cells are cultured in the presence of a test substance. The  
CC levels of LAP mRNA are then measured to determine whether the substance  
CC is an up-regulator. LAP is used to treat microbial infections of the  
CC epithelium (or of those that extend to deeper tissues) e.g. in  
CC immunodeficient states (AIDS), cystic fibrosis, gum disease, wounds,  
CC burns and pneumonia.  
SQ Sequence 44 AA;

Query Match 100.0%; Score 19; DB 1; Length 44;  
Best Local Similarity 25.0%; Pred. No. 1.96e+03;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 20 PIRCPGSM 27  
| |  
QY 2 PXXXXXXM 9

RESULT 5  
ID R88973 standard; Protein; 58 AA.  
AC R88973;  
DT 18-APR-1996 (first entry)  
DE Mutant serine protease inhibitor IV-57C.  
KW Serine protease; inhibitor; factor 8a; factor 11a; plasma kallikrein;  
KW plasmin; thrombolysis; inflammation; septic shock; hypotension; ARDS;  
KW DiC; cardiopulmonary bypass surgery.  
OS Synthetic.  
FH Key Location/Qualifiers  
FT peptide 1..10 /note= "a preferred specific N-terminal peptide,  
FT where the peptide comprises 5-250 amino acids  
FT R88935" and at least one residue is C (see R88928-  
FT 20..33 R88935)"  
FT peptide /note= "a preferred specific internal peptide,  
FT where the peptide comprises 14 amino acids  
FT R88944" and at least one residue is C (see R88936-  
FT 35..37 R88944)"  
FT peptide



```
FT region 25..33 /label= beta-strand_B
FT region 34..42 /label= loop_BC
FT region 43..48 /label= beta-strand_C
FT region 49..57 /label= loop_CD
FT region 58..65 /label= beta-strand_D
FT region 66..67 /label= loop_DE
FT region 68..78 /label= beta-strand_E
FT region 79..86 /label= loop_EF
FT region 87..94 /label= beta-strand_F
FT region 95..100 /label= loop_FG
FT region 101..105 /label= beta-strand_G
FT misc_difference 59 /label= mutation
FT /note= "K -> A"
PN WO9304173-A.
PD 04-MAR-1993.
PF 14-AUG-1992; U06860.
PR 14-AUG-1991; US-744768.
PR 07-MAY-1992; US-879495.
PA (GETH ) GENENTECH INC.
PI Jardieu PM, Presta LG;
DR WPI: 93-094004/11.
PT Polypeptide(s) binding to specific Fc epsilon receptors - act as
PT IGE antagonists; useful for treating and preventing IGE-mediated
PT disorders e.g. allergies
PS Disclosure; Page 73; 113pp; English.
CC IGE mutants were prepd. to evaluate their effect on binding to
CC anti-IGE, esp. MaE11, and to Fc epsilon RI and Fc epsilon RII.
CC Some of the mutants were designed to substitute for a specific
CC amino acid residue another residue with either similar or very
CC different charge or size.
CC Mutant 36 shows +ve binding to Fc epsilon RI and Fc epsilon RII.
SQ Sequence 110 AA;

Query Match 100.0%; Score 19; DB 1; Length 110;
Best Local Similarity 25.0%; Pred. No. 1.96e+03;
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 94 PHLPRALM 101
QY 2 PXXXXXXM 9

RESULT 9
ID R32035 standard; Protein; 110 AA.
AC R32035;
DT 05-JUL-1993 (first entry)
DE Variant IGE - mutant Emut 35.
KW High affinity; FCEH; low affinity; FCEI; Padlan;
KW IGE receptor; Fc; IgG1.
OS Homo sapiens.
FH Key Location/Qualifiers
FT region 7..12 /label= beta-strand_A
FT region 13..24 /label= loop_AB
FT region 25..33 /label= beta-strand_B
FT region 34..42 /label= loop_BC
FT region 43..48 /label= beta-strand_C
```

```
FT region 49..57 /label= loop_CD
FT region 58..65 /label= beta-strand_D
FT region 66..67 /label= loop_DE
FT region 68..78 /label= beta-strand_E
FT region 79..86 /label= loop_EF
FT region 87..94 /label= beta-strand_F
FT region 95..100 /label= loop_FG
FT region 101..105 /label= beta-strand_G
FT misc_difference 58 /label= mutation
FT /note= "R -> A"
PN WO9304173-A.
PD 04-MAR-1993.
PF 14-AUG-1992; U06860.
PR 14-AUG-1991; US-744768.
PR 07-MAY-1992; US-879495.
PA (GETH ) GENENTECH INC.
PI Jardieu PM, Presta LG;
DR WPI: 93-094004/11.
PT Polypeptide(s) binding to specific Fc epsilon receptors - act as
PT IGE antagonists; useful for treating and preventing IGE-mediated
PT disorders e.g. allergies
PS Disclosure; Page 73; 113pp; English.
CC IGE mutants were prepd. to evaluate their effect on binding to
CC anti-IGE, esp. MaE11, and to Fc epsilon RI and Fc epsilon RII.
CC Some of the mutants were designed to substitute for a specific
CC amino acid residue another residue with either similar or very
CC different charge or size.
CC Mutant 35 shows +ve binding to Fc epsilon RI and Fc epsilon RII.
SQ Sequence 110 AA;

Query Match 100.0%; Score 19; DB 1; Length 110;
Best Local Similarity 25.0%; Pred. No. 1.96e+03;
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 94 PHLPRALM 101
QY 2 PXXXXXXM 9

RESULT 10
ID R22037 standard; Protein; 167 AA.
AC R22037;
DT 23-JUL-1992 (first entry)
DE Lymphotoxin protein.
KW Recombinant; LT.
OS Synthetic.
PN J04045789-A.
PD 14-FEB-1992.
PF 14-JUN-1990; 153878.
PR 14-JUN-1990; JP-153878.
PA (TSUR ) TSUMURA & CO.
DR WPI: 92-101940/13.
DR N-PSDB: Q22228.
PT Lymphotoxin expression vector - prepd. from host microbe
PS Disclosure; Fig 4; 15pp; Japanese.
CC The head portion of a LT structural gene (JP234800/89) was synthesised.
CC DNA fragments were recovered and an oligonucleotide contg. a
CC termination codon and a HindIII site was prepd. Bluescript M13SKII
CC was cleaved by EcoRI/HindIII to give plasmid DNA of 2.9 kb. A
CC plasmid contg. a trp promoter and the LT structural gene was
CC constructed (pBtrpLT). The plasmid was expressed and the protein
CC prod. purified by polyethylene imine treatment, dialysis, heat
CC treatment, DEAE Sepharose Fast Flow and CM Sepharose Fast Flow column
CC chromatography, yielding a protein of 95 percent purity and has high
```

CC homology to the natural LT. The method allows greater prodn. of LT  
 CC than by conventional methods.  
 SQ Sequence 167 AA;

Query Match 100.0%; Score 19; DB 1; Length 167;  
 Best Local Similarity 25.0%; Pred. No. 1.96e+03;  
 Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 109 PLLSSQKM 116  
 |  
 QY 2 PXXXXXXM 9

RESULT 11  
 ID R40945 standard; Protein; 194 AA.  
 AC R40945;  
 DT 07-MAR-1994 (first entry)  
 DE Plasminostreptin gene in C-POT type plasmid.  
 KW Detergent; protease; protease inhibitor; plasminostreptin;  
 KW subtilisin; Streptomyces antifibrinolyticus.  
 OS Synthetic.  
 FH Key Location/Qualifiers  
 FT Signal\_peptide 1..85  
 FT /tag= b  
 FT /note= "MF alpha-1 leader sequence"  
 FT mat\_peptide 86..194  
 FT /tag= c  
 FT /label= plasminostreptin

PN W09317086-A.  
 PD 02-SEP-1993.  
 PF 23-FEB-1993; DK0063.  
 PR 25-FEB-1992; DK-000236.  
 PA (NOVO ) NOVO-NORDISK AS.  
 PI Clausen IG, Halkier T, Nielsen LK;  
 DR WPI; 93-288393/36.  
 DR N-PSDB; Q48653.  
 PT Detergent compsn. comprising protease and plasminostreptin as  
 PT protease inhibitor - is useful for liq. detergents contg.  
 PT Oxidising agent requiring enzyme stability  
 PS Claim 1; Page 17; 28pp; English.  
 CC The C-POT type plasmid contains the Schizosaccharomyces pombe triose  
 CC phosphate isomerase gene (POT) for the purpose of plasmid  
 CC stabilization and the Saccharomyces cerevisiae triose phosphate  
 CC isomerase promoter and terminator (P-TPI and T-TPI). The fragment  
 CC encoding a signal/leader/insulin precursor sequence is replaced by  
 CC a fragment encoding the MF-alpha 1 leader fused to the plasminostreptin  
 CC sequence (Q35503).  
 CC The detergent comprises a protease protease (pref. subtilisin  
 CC from Bacillus) and a reversible protease inhibitor,  
 CC (plasminostreptin from Streptomyces antifibrinolyticus) The  
 CC inhibitor is used to stabilise the protease in the detergent.  
 CC The protease is often so strongly bound to the inhibitor that  
 CC little protease activity is released when the inhibitor is diluted  
 CC for use. A plasminostreptin variant exhibiting a weaker binding  
 CC to the protease is preferred for use in detergents.  
 SQ Sequence 194 AA;

Query Match 100.0%; Score 19; DB 1; Length 194;  
 Best Local Similarity 25.0%; Pred. No. 1.96e+03;  
 Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 90 PSALVLTM 97  
 |  
 QY 2 PXXXXXXM 9

RESULT 12  
 ID R60361 standard; Protein; 199 AA.  
 AC R60361;  
 DT 28-FEB-1995 (first entry)  
 DE MnS2:MnSOD variant.  
 KW Manganese superoxide dismutase; MnSOD; oxygen; cosmetic;  
 KW mutation; arthritis; variant; enzyme; isoelectric point;

KW inflammation; cancer; premature retinopathy; hypertension;  
 KW diabetes; ss.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT cds 7..606  
 FT /tag= a  
 FT misc\_difference 406..408  
 FT /tag= b  
 FT /transl\_except= pos:406..408, aa:Gly

PN W09414950-A.  
 PD 07-JUL-1994.  
 PF 28-DEC-1993; J01917.  
 PR 28-DEC-1992; JP-359959.  
 PA (SIIT-) SII TECHNORESEARCH INC.  
 PI Katsuta K, Kondo M;  
 DR WPI; 94-234688/28.  
 DR N-PSDB; Q70432.  
 PT Human variant manganese super-oxide dismutase with aminoacid  
 PT mutation in non-essential regions - for treatment of arthritis,  
 PT active oxygen related disorders and for cosmetics  
 PS Example 13; Page 19-20; 34pp; Japanese.  
 CC Human variant manganese superoxide dismutase (Mn-SOD) comprises  
 CC an amino acid substitution in a region which does not affect enzyme  
 CC activity, of a positively charged amino acid residue, which has the  
 CC effect of raising the isoelectric point of the peptide. These  
 CC Mn-SOD variants are useful as medical preps. for the treatment  
 CC of human arthritis and diseases caused by active oxygen, and as  
 CC cosmetic preps. They are also useful for treatment of inflammation,  
 CC cancer, premature retinopathy, hypertension and diabetes.  
 CC Examples of variant Mn-SODs are given in Q70431-32.  
 SQ Sequence 199 AA;

Query Match 100.0%; Score 19; DB 1; Length 199;  
 Best Local Similarity 25.0%; Pred. No. 1.96e+03;  
 Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 17 PHINAQIM 24  
 |  
 QY 2 PXXXXXXM 9

RESULT 13  
 ID R82934 standard; Protein; 230 AA.  
 AC R82934;  
 DT 26-FEB-1996 (first entry)  
 DE Interleukin 4 component common to the IL-2 receptor gamma chain.  
 DE Interleukin-4; IL-4; gamma chain component; immunosuppressants;  
 KW anti-allergy agent; signal transduction inhibitor; autoimmune;  
 KW disease; anti-inflammatories; anaphylactic shock; bronchial asthma;  
 KW Interleukin-2; IL-2; atopic dermatitis; urticaria.  
 OS Homo sapiens.  
 PN J07149662-A.  
 PD 13-JUN-1995.  
 PF 07-SEP-1994; 213706.  
 PR 08-SEP-1993; JP-223574.  
 PA (AJIN ) AJINOMOTO KK.  
 PA (SUGA/) SUGAMURA K.  
 DR WPI; 95-243601/32.  
 DR N-PSDB; T04952.  
 PT Novel interleukin-4 receptor monoclonal antibodies inhibit signal  
 PT transmission - useful as immunosuppressants and anti-allergy agents.  
 PS Example 1; Page 9; 11pp; Japanese.  
 CC T04952 encodes R82934 a component of the IL-4 receptor common to  
 CC the IL-2 receptor gamma chain molecule, which was used to generate  
 CC anti-IL-4 receptor monoclonal antibodies (mAbs). The mAbs (IL-4  
 CC signal transduction inhibitors) can be used as immunosuppressants  
 CC and anti-allergy agents, for the treatment of autoimmune and chronic  
 CC inflammatory diseases, e.g. anaphylactic shock, bronchial asthma,  
 CC atopic dermatitis and urticaria.  
 SQ Sequence 230 AA;

Query Match 100.0%; Score 19; DB 1; Length 230;  
 Best Local Similarity 25.0%; Pred. No. 1.96e+03;

Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 117 PRQATQM 124  
QY 2 PXXXXXXM 9

## RESULT 14

ID R11178 standard; Protein; 257 AA.  
AC R11178; 1991 (first entry)  
DE Plasmid pBTA735-encoded TraTp-LHRH fusion protein.  
KW TraTp protein; Leutinizing hormone releasing hormone; fusion protein;  
KW immunological castration.  
FH Key Location/Qualifiers  
FT peptide 1..20  
FT /label= TraTp signal  
FT peptide 103..112  
FT /label= LHRH analogue  
FT protein 21..257  
FT /label= TraTp-LHRH fusion  
PN W09102799-A.  
PD 07-MAR-1991.  
PF 24-AUG-1990; AU0373.  
PR 25-AUG-1989; AU-005979.  
PA (BIOT-) BIOTECHN AUST PTY L.  
PI Russell-Jones GJ, Stewart AG, Tsonis CG;  
DR WPI; 91-087282/12.  
DR N-PSDB; Q10999.  
PT Fusion proteins comprising LHRH analogue and TraTp (analogue) -  
PT useful in vaccine for inhibition or control of reproduction in  
PT vertebrates, esp. domestic animals  
PS Example 1: Fig 2 and 3: 53pp; English.  
CC Plasmid pBTA735 is a TraTp-LHRH analogue fusion in which the LHRH  
CC analogue has been inserted between amino acids 101 and 102 of TraTp  
CC (Ogata R.T. et al., (1982) J.Bacteriol. 151:819-827). The synthetic  
CC LHRH analogue was inserted via a linker which provides a unique new  
CC SmaI site located between codons such that the LHRH is inserted in  
CC frame. The fusion protein can be used in vaccines for the  
CC inhibition or control of reproduction in vertebrates, where the  
CC TraTp acts as an adjuvant.  
CC See also Q10995, Q10997-8, Q11000, Q11014-Q11021.  
SQ Sequence 257 AA;

Query Match 100.0%; Score 19; DB 1; Length 257;  
Best Local Similarity 25.0%; Pred.No. 1.96e+03;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 111 PGSSDKM 118  
QY 2 PXXXXXXM 9

## RESULT 15

ID R66901 standard; Protein; 269 AA.  
AC R66901;  
DT 03-JUL-1995 (first entry)  
DE M. tuberculosis inhA.  
KW Isoniazid; isonicotinic acid hydrazide; INH; inhA gene; vaccine.  
OS Mycobacterium tuberculosis.  
PN W09426765-A.  
PD 24-NOV-1994.  
PF 13-MAY-1994; U05398.  
PR 13-MAY-1993; NZ-247620.  
PR 14-MAY-1993; US-062409.  
PR 31-MAR-1994; US-221742.  
PA (AGRE-) AGRESEARCH.  
PA (BANE/) BANERJEE A.  
PA (COLL/) COLLINS D.  
PA (JACO/) JACOBS W R.  
PA (YESH) UNIV YESHIVA EINSTEIN COLLEGE.  
PA (WILS/) WILSON T M.  
PI Banerjee A, Collins D, De LISIE GW, Jacobs WR, Wilson TM;

DR WPI; 95-006691/01.  
DR N-PSDB; Q75518.  
PT Polynucleotide(s) determining mycobacterial resistance to  
PT isoniazid, useful in diagnosis, treatment and prevention of  
PT mycobacterial infection, e.g. tuberculosis.  
PS Disclosure; Fig. 4; 104pp; English.  
CC The gene from Mycobacterium tuberculosis encoding inhA (R66901),  
CC the target of action for isoniazid, was identified, isolated,  
CC cloned and sequenced (Q75518). Mutant inhA genes have been used  
CC for recombinant vaccine development.  
SQ Sequence 269 AA;

Query Match 100.0%; Score 19; DB 1; Length 269;  
Best Local Similarity 25.0%; Pred.No. 1.96e+03;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 131 PRALLPTM 138  
QY 2 PXXXXXXM 9

Search completed: Sat Apr 15 02:17:29 2000  
Job time : 37 secs.



**HIS PAGE BLANK (USPTO)**

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL; M60792; AAA72981.1; -;  
DR FLXBASE; FBgn0012535; DmicVAdh.  
DR PROSITE; PS00061; ADH\_SHORT; 1.  
DR PFAM; PF00106; adh\_short; 1.  
DR PFAM; PF00663; adh\_short\_C; 1.  
KW Oxidoreductase; NAD. 0 BY SIMILARITY.  
FT INIT\_MET 0 0 NAD (BY SIMILARITY).  
FT NP\_BIND 9 32 NAD (BY SIMILARITY).  
FT ACT\_SITE 150 150 BY SIMILARITY.  
SQ SEQUENCE 253 AA; 27345 MW; A3B9DD6A CRC32;  
  
Query Match 100.0%; Score 16; DB 1; Length 253;  
Best Local Similarity 25.0%; Pred. No. 2.99e+03;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
  
Db 60 PYDVTVP 67  
|  
QY 2 PXXXXXL 9  
  
RESULT 15  
ID CHMU\_YEAST STANDARD; PRT; 256 AA.  
AC P32178;  
DT 01-OCT-1993 (Rel. 27, Created)  
DT 01-OCT-1993 (Rel. 27, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE CHORISMATE MUTASE (EC 5.4.99.5) (CM).  
GN ARO7 OR OSM2 OR YPR060C OR YP9499.15C.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;  
OC Saccharomycetaceae; Saccharomyces.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-X2180;  
RX MEDLINE; 89155418.  
RA SCHMIDHEINI T., SPERISEN P., PARAVICINI G., HUETTER R., BRAUS G.H.;  
RT "A single point mutation results in a constitutively activated and  
feedback-resistant chorismate mutase of Saccharomyces cerevisiae.";  
RL J. Bacteriol. 171:1245-1253(1989).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-S288C / AB972;  
RX MEDLINE; 95062155.  
RA XUE Y., LIPSCOMB W.N., GRAF R., SCHNAPPAUF G., BRAUS G.;  
RT "The crystal structure of allosteric chorismate mutase at 2.2-A  
resolution.";  
RL Proc. Natl. Acad. Sci. U.S.A. 91:10814-10818(1994).  
RN [4]  
RP X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS).  
RX MEDLINE; 96194968.  
RA STRAETER N., HARANSSON K., SCHNAPPAUF G., BRAUS G., LIPSCOMB W.N.;  
RT "Crystal structure of the T state of allosteric yeast chorismate  
mutase and comparison with the R state.";  
RL Proc. Natl. Acad. Sci. U.S.A. 93:3330-3334(1996).  
RN [5]  
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).  
RX MEDLINE; 98046093.  
RA STRAETER N., SCHNAPPAUF G., BRAUS G., LIPSCOMB W.N.;  
RT "Mechanisms of catalysis and allosteric regulation of yeast  
chorismate mutase from crystal structures.";

RL Structure 5:1437-1452(1997).  
CC -|- CATALYTIC ACTIVITY: CHORISMATE - PREPHENATE.  
CC -|- ENZYME REGULATION: NEEDS TRYPTOPHAN FOR ACTIVATION AND TYROSINE IS  
CC A STRONG INHIBITOR. ALLOSTERICALLY REGULATED.  
CC -|- PATHWAY: BRANCH POINT OF THE BIOSYNTHETIC PATHWAY LEADING TO THE  
CC THREE AROMATIC AMINO ACIDS, PHENYLALANINE, TYROSINE, & TRYPTOPHAN.  
CC -|- SUBUNIT: HOMODIMER.  
CC -|- SIMILARITY: TO A THALIANA CHORISMATE MUTASE.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL; M24517; AAB59309.1; -;  
DR EMBL; Z49219; CAA89177.1; -;  
DR EMBL; Z71255; CAA95004.1; -;  
DR PIR; A45921; A45921.  
DR PDB; 1CSM; 15-SEP-95.  
DR PDB; 2CSM; 23-DEC-96.  
DR PDB; 3CSM; 14-JAN-98.  
DR PDB; 4CSM; 14-JAN-98.  
DR PDB; 5CSM; 14-JAN-98.  
DR SGD; L0000120; ARO7.  
KW Aromatic amino acid biosynthesis; Isomerase; 3D-structure;  
KW Allosteric enzyme.  
FT VARIANT 226 226 T -> I (CONSTITUTIVELY ACTIVATED AND  
FEEDBACK-RESISTANT).  
SQ SEQUENCE 256 AA; 29747 MW; B8B0BC0A CRC32;  
  
Query Match 100.0%; Score 16; DB 1; Length 256;  
Best Local Similarity 25.0%; Pred. No. 2.99e+03;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
  
Db 53 PNFKGSFL 60  
|  
QY 2 PXXXXXL 9  
  
Search completed: Sat Apr 15 02:11:56 2000  
Job time : 39 secs.

CC FA

```
SQ SEQUENCE 221 AA; 25307 MW; 46DFCE27 CRC32;
Query Match 100.0%; Score 16; DB 1; Length 221;
Best Local Similarity 25.0%; Pred.No. 2.99e+03;
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 146 PSPKSTL 153
|
QY 2 PXXXXXXL 9

RESULT 9
ID ATP6_BALMU STANDARD; PRT; 226 AA.
AC P41291;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE ATP SYNTHASE A CHAIN (EC 3.6.1.34) (PROTEIN 6).
GN MTATP6 OR ATP6.
OS Balaenoptera musculus (Blue whale).
OG Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Cetartiodactyla; Cetacea; Mysticeti; Balaenopteridae;
OC Balaenoptera.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 94141932.
RA ARNASON U., GULLBERG A.;
RT "Comparison between the complete mtDNA sequences of the blue and the
J. Mol. Evol. 37:312-322(1993).
CC -!- FUNCTION: KEY COMPONENT OF THE PROTON CHANNEL; IT MAY PLAY A
DIRECT ROLE IN THE TRANSLLOCATION OF PROTONS ACROSS THE MEMBRANE.
CC -!- SUBUNIT: F-TYPE ATPASES HAVE 2 COMPONENTS, CF(1) - THE CATALYTIC
CORE - AND CF(0) - THE MEMBRANE PROTON CHANNEL. CF(1) HAS FIVE
SUBUNITS: ALPHA(3), BETA(3), GAMMA(1), DELTA(1), EPSILON(1). CF(0)
HAS THREE MAIN SUBUNITS: A, B AND C.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -!- SIMILARITY: BELONGS TO THE ATPASE A CHAIN FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
CC DR EMBL; X72204; CAA51000.1; -
DR PIR; S41825; S41825.
DR PROSITE; PS00449; ATPASE_A; 1.
DR PFAM; PF00119; ATP-synt_A; 1.
KW Hydrogen ion transport; CF(0); Mitochondrion; Transmembrane.
SQ SEQUENCE 226 AA; 24967 MW; 6B1DD99F CRC32;
Query Match 100.0%; Score 16; DB 1; Length 226;
Best Local Similarity 25.0%; Pred.No. 2.99e+03;
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 8 PFMIPVYL 15
|
QY 2 PXXXXXXL 9

RESULT 10
ID CRTA_RHOCA STANDARD; PRT; 241 AA.
AC P17055;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 01-MAY-1992 (Rel. 22, Last annotation update)
DE SPHEROIDENE MONOOXYGENASE (EC 1.-.-.-).
GN CRTA.
OS Rhodobacter capsulatus (Rhodospseudomonas capsulata).
-----
OC Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;
OC Rhodobacter.
RN [1]
RP SEQUENCE FROM N.A.
RA BURKE D.H., ALBERTI M., ARMSTRONG G.A., HEARST J.E.;
RL Submitted (NOV-1991) to the EMBL/GenBank/DBJ databases.
RN [2]
RP PRELIMINARY SEQUENCE FROM N.A.
RC STRAIN-SB1003, AND BEC404;
RX MEDLINE; 89313663.
RA ARMSTRONG G.A., ALBERTI M., LEACH F., HEARST J.E.;
RT "Nucleotide sequence, organization, and nature of the protein
products of the carotenoid biosynthesis gene cluster of Rhodobacter
capsulatus.";
RL Mol. Gen. Genet. 216:254-268(1989).
CC -!- PATHWAY: CAROTENOID AND CHLOROPHYLL BIOSYNTHESIS.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
CC DR EMBL; Z11165; CAA7539.1; -
DR EMBL; X52291; CAA36532.1; -
DR PIR; S04401; S04401.
DR PIR; S17822; S17822.
KW Photosynthesis; Chlorophyll biosynthesis; Carotenoid biosynthesis;
KW Oxidoreductase.
SQ SEQUENCE 241 AA; 27004 MW; 59085F33 CRC32;
Query Match 100.0%; Score 16; DB 1; Length 241;
Best Local Similarity 25.0%; Pred.No. 2.99e+03;
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 35 PRVKFVKL 42
|
QY 2 PXXXXXXL 9

RESULT 11
ID CD8A_BOVIN STANDARD; PRT; 242 AA.
AC P31783;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE T-CELL SURFACE GLYCOPROTEIN CD8 ALPHA CHAIN PRECURSOR.
GN CD8A.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=THYMUS;
RX MEDLINE; 92332098.
RA LALOR P., BUCCI C., FORNARO M., RATTAZZI M.C., NAKAUCHI H.,
RA HERZENBERG L.A., ALBERTI S.;
RT "Molecular cloning, reconstruction and expression of the gene
encoding the alpha-chain of the bovine CD8 -- definition of three
peptide regions conserved across species.";
RL Immunology 76:95-102(1992).
CC -!- FUNCTION: IDENTIFIES CYTOTOXIC/SUPPRESSOR T-CELLS THAT INTERACT
WITH MHC CLASS I BEARING TARGETS. CD8 IS THOUGHT TO PLAY A ROLE IN
THE PROCESS OF T-CELL MEDIATED KILLING. CD8 ALPHA CHAINS BINDS TO
CLASS MHC MOLECULES ALPHA-3 DOMAINS.
CC -!- SUBUNIT: IN GENERAL HETERODIMER OF AN ALPHA AND A BETA CHAIN
LINKED BY TWO DISULFIDE BONDS. CAN ALSO FORMS HOMODIMERS.
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.
-----
```

SEQUENCE OF 19-192 FROM N.A.  
 RP MEDLINE: 91169528.  
 RA NAKAHORI Y., TAKENAKA O., NAKAGOME Y.: 'amelogenin'";  
 RT "A human X-Y homologous region encodes  
 RL Genomics 9:264-269(1991).  
 CC -1- FUNCTION: PLAYS A ROLE IN BIOMINERALIZATION. SEEM TO REGULATE THE  
 CC FORMATION OF CRYSTALLITES DURING THE SECRETORY STAGE OF TOOTH  
 CC ENAMEL DEVELOPMENT. THOUGHT TO PLAY A MAJOR ROLE IN THE STRUCTURAL  
 CC ORGANIZATION AND MINERALIZATION OF DEVELOPING ENAMEL.  
 CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR MATRIX.  
 CC -1- ALTERNATIVE PRODUCTS: VARIOUS FORMS ARE PRODUCED BY ALTERNATIVE  
 CC SPLICING.  
 CC -1- DEVELOPMENTAL STAGE: TRANSIENTLY BUT ABUNDANTLY EXPRESSED BY  
 CC AMELOBLASTS DURING TOOTH DEVELOPMENT. AMELOGENIN IS THE  
 CC PREDOMINANT PROTEIN IN DEVELOPING DENTAL ENAMEL.  
 CC -1- MISCELLANEOUS: THIS ISOFORM IS ENCODED BY THE GENE ON CHROMOSOME  
 CC Y.  
 CC -1- SIMILARITY: BELONGS TO THE AMELOGENIN FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: M86933; AAA51718.1; -;  
 DR EMBL: X14439; CAA32612.1; -;  
 DR EMBL: M55419; AAG62827.1; -;  
 DR PIR: F41816; F41816.  
 DR MIM: 410000; -;  
 KW Extracellular matrix; Phosphorylation; Enamel; Repeat; Signal;  
 KW Alternative splicing.  
 FT SIGNAL 16 BY SIMILARITY.  
 FT CHAIN 17 192 AMELOGENIN, Y ISOFORM.  
 FT MOD\_RES 32 32 PHOSPHORYLATION (BY SIMILARITY).  
 FT CONFLICT 192 192 D -> VSTP (IN REF. 2).  
 SQ SEQUENCE 192 AA; 21730 MW; 688D9D6F CRC32;  
 Query Match 100.0%; Score 16; DB 1; Length 192;  
 Best Local Similarity 25.0%; Pred. No. 2.99e+03;  
 Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
 Db 94 PRVROQAL 101  
 QY 2 PXXXXXXL 9  
 RESULT 7  
 ID CYSR\_SYNY3 STANDARD; PRT; 205 AA.  
 AC Q5854;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE REGULATORY PROTEIN CYSR HOMOLOG.  
 GN CYSR OR SLL0594.  
 OS Synechocystis sp. (strain PCC 6803).  
 OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 96127529.  
 RA KANEKO T., TANAKA A., SATO S., KOTANI H., SAZUKA T., MIYAJIMA N.,  
 RA SUGIURA M., TABATA S.;  
 RT "Sequence analysis of the genome of the unicellular cyanobacterium  
 RT Synechocystis sp. strain PCC6803. I. Sequence features in the 1 Mb  
 RT region from map positions 64% to 92% of the genome.";  
 RL DNA Res. 2:153-166(1995).  
 CC -1- FUNCTION: PROBABLY REGULATES THE EXPRESSION OF GENES FROM THE  
 CC SULFATE PERMEASE COMPLEX (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (POTENTIAL).  
 CC -1- SIMILARITY: BELONGS TO THE CRP/FNR FAMILY OF TRANSCRIPTIONAL  
 CC REGULATORS.

CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: D64004; BAA10610.1; -;  
 DR PROSITE: PS00042; HTR\_CRP\_FAMILY; 1.  
 DR PFAM: PF00325; crp; 1.  
 KW Sulfate transport; Transport; Transcription regulation; DNA-binding.  
 FT DNA\_BIND 164 183 H-T-H MOTIF (POTENTIAL).  
 SQ SEQUENCE 205 AA; 22656 MW; ADE04764 CRC32;  
 Query Match 100.0%; Score 16; DB 1; Length 205;  
 Best Local Similarity 25.0%; Pred. No. 2.99e+03;  
 Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
 Db 151 PGTGMRL 158  
 QY 2 PXXXXXXL 9  
 RESULT 8  
 ID CD28\_RABIT STANDARD; PRT; 221 AA.  
 AC P42069;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 01-OCT-1996 (Rel. 34, Last annotation update)  
 DE T-CELL-SPECIFIC SURFACE GLYCOPROTEIN CD28 PRECURSOR.  
 GN CD28.  
 OS Oryctolagus cuniculus (Rabbit).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Lagomorpha; Leporidae; Oryctolagus.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-B/J X CHBB.HM;  
 RX MEDLINE: 95369849.  
 RA ISONO T., SETO A.;  
 RT "Cloning and sequencing of the rabbit gene encoding T-cell  
 RT costimulatory molecules.";  
 RL Immunogenetics 42:217-220(1995).  
 CC -1- FUNCTION: POSSIBLY INVOLVED IN T-CELL ACTIVATION. BINDS TO B7-1  
 CC AND B7-2 (B70) (BY SIMILARITY).  
 CC -1- SUBUNIT: HOMODIMER, LINKED BY A DISULFIDE BOND (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.  
 CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. STRONGEST  
 CC SIMILARITY TO CTLA-4.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: D49841; BAA08641.1; -;  
 KW Immunoglobulin domain; T-cell; Glycoprotein; Signal; Transmembrane.  
 FT SIGNAL 1 19 BY SIMILARITY.  
 FT CHAIN 20 221 T-CELL-SPECIFIC SURFACE GLYCOPROTEIN  
 FT CD28.  
 FT DOMAIN 20 150 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 151 177 POTENTIAL.  
 FT DOMAIN 178 221 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 29 138 IG-LIKE V-TYPE DOMAIN.  
 FT CARBOHYD 38 38 POTENTIAL.  
 FT CARBOHYD 72 72 POTENTIAL.  
 FT CARBOHYD 93 93 POTENTIAL.  
 FT CARBOHYD 106 106 POTENTIAL.  
 FT CARBOHYD 130 130 POTENTIAL.

RP SEQUENCE FROM N.A.  
RX MEDLINE: 87203384.  
RA GOPE M.L., KEINADEN R.A., KRISTO P.A., CONNEELY O.M., BEATTIE W.G.,  
RA ZARUCKI-SCHULZ T., O'MALLEY B.W., KULOMAA M.S.:  
RT "Molecular cloning of the chicken avidin cDNA."  
RL Nucleic Acids Res. 15:3595-3606(1987).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 90355928.  
RA CHANDRA G., GRAY J.G.:  
RT "Cloning and expression of avidin in Escherichia coli."  
RL Meth. Enzymol. 184:70-79(1990).  
RN [3]  
RP SEQUENCE FROM N.A.  
RX STRAIN-WHITE LECHORN; TISSUE-OVIDUCT;  
RX MEDLINE: 95394357.  
RA WALLIN M.J., LAUKKANEN M.O., KULOMAA M.S.:  
RT "Cloning and sequencing of the chicken egg-white avidin-encoding gene  
RT and its relationship with the avidin-related genes Avrl-Avr5."  
RL Gene 161:205-209(1995).  
RN [4]  
RP SEQUENCE OF 25-152.  
RX MEDLINE: 71107558.  
RA DELANGE R.J., HUANG T.-S.:  
RT "Egg white avidin. 3. Sequence of the 78-residue middle cyanogen  
RT bromide peptide. Complete amino acid sequence of the protein  
RT subunit."  
RL J. Biol. Chem. 246:698-709(1971).  
RN [5]  
RP IMPORTANCE OF TYR IN BIOTIN-BINDING.  
RX MEDLINE: 90351377.  
RA GITLIN G., BAYER E.A., WILCHEK M.:  
RT "Studies on the biotin-binding sites of avidin and streptavidin.  
RT Tyrosine residues are involved in the binding site."  
RL Biochem. J. 269:527-530(1990).  
RN [6]  
RP BIOTIN-BINDING STUDIES.  
RX MEDLINE: 91378911.  
RA HILLER Y., BAYER E.A., WILCHEK M.:  
RT "Studies on the biotin-binding site of avidin. Minimized fragments  
RT that bind biotin."  
RL Biochem. J. 278:573-585(1991).  
RN [7]  
RP X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS).  
RX MEDLINE: 93281699.  
RA LYNNAH O., BAYER E.A., WILCHEK M., SUSSMAN J.L.:  
RT "Three-dimensional structures of avidin and the avidin-biotin  
RT complex."  
RL Proc. Natl. Acad. Sci. U.S.A. 90:5076-5080(1993).  
RN [8]  
RP X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS).  
RX MEDLINE: 93294833.  
RA PUGLIESE L., CODA A., MALCOVATI M., BOLOGNESI M.:  
RT "Three-dimensional structure of the tetragonal crystal form of  
RT egg-white avidin in its functional complex with biotin at 2.7-A  
RT resolution."  
RL J. Mol. Biol. 231:698-710(1993).  
RN [9]  
RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS).  
RX MEDLINE: 93294833.  
RA ROSANO C., AROSIO P., BOLOGNESI M.:  
RT Submitted (MAR-1998) to the PDB data bank.  
RL  
CC -!- FUNCTION: THE BIOLOGICAL FUNCTION OF AVIDIN IS NOT KNOWN. FORMS A  
CC STRONG NON-COVALENT SPECIFIC COMPLEX WITH BIOTIN (ONE MOLECULE OF  
CC BIOTIN PER SUBUNIT OF AVIDIN).  
CC -!- SUBUNIT: HOMOTETRAMER.  
CC -!- TISSUE SPECIFICITY: SYNTHESIZED IN HEN OVIDUCT AND CONCENTRATED IN  
CC EGG WHITE (WHERE IT REPRESENTS 0.05% OF THE TOTAL PROTEIN).  
CC -!- SIMILARITY: BELONGS TO THE AVIDIN/STREPTAVIDIN FAMILY.  
CC -!- DATABASE: NAME-Worthington enzyme manual;  
CC WWW="http://www.worthington-biochem.com/manual/AV/AV.html".  
CC -!- DATABASE: NAME-Prozyme technical fact sheet;  
CC WWW="http://www.prozyme.com/technical/av10data.html".  
RN [2]

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; X05343; CAA28954.1; -;  
CC EMBL; L27818; AAB59733.1; -;  
CC PIR; A03160; VICH.  
CC PIR; A27518; A27518.  
CC PIR; S11540; S11540.  
CC PDB; 2AVI; 15-JUL-93.  
CC PDB; 1AVD; 31-JAN-94.  
CC PDB; 1AVE; 31-JAN-94.  
CC PDB; 1RAV; 15-JUL-98.  
CC PDB; 2CAM; 15-JUL-98.  
CC PROSITE; PS00577; AVIDIN; 1.  
CC PFAM; PF01382; Avidin; 1.  
KW Glycoprotein; Signal; Biotin; 3D-structure.  
FT SIGNAL 1 24  
FT CHAIN 25 152 AVIDIN.  
FT DISULFID 28 107  
FT CARBOHYD 41 41  
FT BINDING 57 57  
FT VARIANT 58 58 INVOLVED IN BIOTIN BINDING.  
FT CONFLICT 22 22 I -> T (IN APPR. 50% OF THE CHAINS).  
FT CONFLICT 77 77 G -> S (IN REF. 3).  
FT STRAND 32 36 E -> Q (IN REF. 2 AND 3).  
FT TURN 37 38  
FT STRAND 41 44  
FT TURN 49 50  
FT STRAND 54 58  
FT STRAND 71 74  
FT STRAND 77 77  
FT TURN 83 84  
FT STRAND 87 93  
FT STRAND 100 109  
FT STRAND 115 124  
FT HELIX 130 135  
FT STRAND 137 146  
SQ SEQUENCE 152 AA; 16769 MW; 570ACC01 CRC32;  
Query Match 100.0%; Score 16; DB 1; Length 152;  
Best Local Similarity 25.0%; Pred No. 2.99e+03;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
Db 7 PILLLLLL 14  
Qy 2 PXXXXXXL 9  
RESULT 6  
ID AMFY HUMAN STANDARD; PRT; 192 AA.  
AC O99218;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 01-NOV-1995 (Rel. 32, Last annotation update)  
DE AMELOGENIN, Y ISOFORM PRECURSOR.  
GN AMELY OR AMGY OR AMGL.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Primates; Catarrhini; Homnidae; Homo.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=TOOTH BUD;  
RX MEDLINE: 92133605.  
RA SALIDO E.C., YEN P.H., KOPRIVNIKAR K., YU L.-C., SHAPIRO L.J.:  
RT "The human enamel protein gene amelogenin is expressed from both the  
RL x and the y chromosomes."  
RL Am. J. Hum. Genet. 50:303-316(1992).  
RN [2]

use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----  
DR EMBL; J01917; AAA92207.1; ALT\_SEQ.  
DR PIR; A03832; DQAD62.  
KW DNA-binding.  
SQ SEQUENCE 145 AA; 16102 MW; 6A0C8E88 CRC32;

Query Match 100.0%; Score 16; DB 1; Length 145;  
Best Local Similarity 25.0%; Pred. No. 2,99e+03;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

DB 71 PLAWRVWL 78  
|  
QY 2 PXXXXXXL 9

## RESULT 4

ID CYNS\_SYNY3 STANDARD; PRT; 149 AA.  
AC Q55367;  
DT 01-NOV-1997 (Rel. 35; Created)  
DT 01-NOV-1997 (Rel. 35; Last sequence update)  
DT 01-NOV-1997 (Rel. 35; Last annotation update)  
DE CYANATE LYASE (EC 4.3.99.1) (CYANATE HYDROLASE) (CYANASE).  
GN CYNS OR SLR0899.  
OS Synchocystis sp. (strain PCC 6803).  
OC Bacteria; Cyanobacteria; Chroococcales; Synchocystis.  
RN [1]

RP SEQUENCE FROM N.A.  
RX MEDLINE; 96127529.  
RA KANEKO T., TANAKA A., SATO S., KOTANI H., SAZUKA T., MIYAJIMA N.,  
RA SUGIURA M., TABATA S.;  
RT "Sequence analysis of the genome of the unicellular cyanobacterium Synchocystis sp. strain PCC6803. I. Sequence features in the 1 Mb region from map positions 64% to 92% of the genome."  
RL DNA Res. 2:153-166(1995).  
CC -|- CATALYTIC ACTIVITY: CYANATE + BICARBONATE = CO(2) + CARBAMATE.

-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----  
DR EMBL; D64003; BAAL0449.1; -  
KW Lyase.  
SQ SEQUENCE 149 AA; 16532 MW; 40BE1E31 CRC32;

Query Match 100.0%; Score 16; DB 1; Length 149;  
Best Local Similarity 25.0%; Pred. No. 2,99e+03;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

DB 80 PVVPTDPL 87  
|  
QY 2 PXXXXXXL 9

## RESULT 5

ID AVID\_CHICK STANDARD; PRT; 152 AA.  
AC P02701; Q91958;  
DT 21-JUL-1986 (Rel. 01; Created)  
DT 01-NOV-1991 (Rel. 20; Last sequence update)  
DT 13-DEC-1999 (Rel. 39; Last annotation update)  
DE AVIDIN PRECURSOR.  
GN AVID.  
OS Gallus gallus (Chicken).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Archosauria; Aves;  
OC Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.  
RN [1]

RESULT 2  
ID B2MG\_CAVPO STANDARD; PRT; 99 AA.  
AC P01886;  
DT 21-JUL-1986 (Rel. 01; Created)  
DT 21-JUL-1986 (Rel. 01; Last sequence update)  
DT 01-OCT-1996 (Rel. 34; Last annotation update)  
DE BETA-2-MICROGLOBULIN.  
GN B2M.  
OS Cavia porcellus (Guinea pig).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Rodentia; Hystricognathi; Caviidae; Cavia.  
RN [1]  
RP SEQUENCE.  
RX MEDLINE; 82057805.  
RA WOLFE P.B., CEBRA J.J.;  
RT "The primary structure of guinea pig beta 2-microglobulin.";  
RL Mol. Immunol. 17:1493-1505(1980).  
CC -|- FUNCTION: BETA-2-MICROGLOBULIN IS THE BETA-CHAIN OF MAJOR HISTOCOMPATIBILITY COMPLEX CLASS I MOLECULES.  
CC -|- SUBCELLULAR LOCATION: EXTRACELLULAR.  
DR PIR; A02181; MGGPB2.  
DR HSP; P01888; LBMG.  
DR PROSITE; PS00290; IG\_MHC; 1.  
DR PFAM; PF00047; Ig; 1.  
KW MHC I.  
FT DISULFID 25 80 BY SIMILARITY.  
SQ SEQUENCE 99 AA; 11417 MW; DEDB12C5 CRC32;

Query Match 100.0%; Score 16; DB 1; Length 99;  
Best Local Similarity 25.0%; Pred. No. 2,99e+03;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

DB 32 PPOIEVEL 39  
|  
QY 2 PXXXXXXL 9

## RESULT 3

ID DNBL\_ADE02 STANDARD; PRT; 145 AA.  
AC P03563;  
DT 21-JUL-1986 (Rel. 01; Created)  
DT 21-JUL-1986 (Rel. 01; Last sequence update)  
DT 01-APR-1990 (Rel. 14; Last annotation update)  
DE PROBABLE DNA-BINDING PROTEIN (AGNOPROTEIN).  
OS Human adenovirus type 2.  
OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 82196890.  
RA VIRTANEN A., ALESTROM P., PERSSON H., KATZE M.G., PETTERSSON U.;  
RT "An adenovirus agnogene."  
RL Nucleic Acids Res. 10:2539-2548(1982).  
[2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 83056843.  
RA GINGERAS T.R., SCIAKY D., GELINAS R.E., BING-DONG J., YEN C.E.,  
RA KELLY M.M., BULLOCK P.A., PARSONS B.L., O'NEILL K.E., ROBERTS R.J.;  
RT "Nucleotide sequences from the adenovirus-2 genome."  
RL J. Biol. Chem. 257:13475-13491(1982).  
RN [3]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 83056844.  
RA ALESTROM P., AKUSJARVI G., PETTERSSON M., PETTERSSON U.;  
RT "DNA sequence analysis of the region encoding the terminal protein and the hypothetical N-gene product of adenovirus type 2.";  
RL J. Biol. Chem. 257:13492-13498(1982).  
CC -|- MISCELLANEOUS: SEE THE ENTRIES FOR DNA-BINDING PROTEINS OF SIMIAN VIRUS 40 AND BK VIRUS.

-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its

\*\*\*\*\*

(TM)

\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.

Copyright (c) 1993-1998 University of Edinburgh, U.K.

Distribution rights by Oxford Molecular Ltd

Mpsrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Sat Apr 15 02:11:17 2000; Maspar time 3.09 seconds

Tabular output not generated. 87.096 Million cell updates/sec

Title: &gt;US-08-452-843-29

Description: (1-9) from US08452843.pap

Perfect Score: 16

Sequence: 1 XPXXXXXXL 9

Scoring table: PAM 150

Gap 15

Searched: 82229 seqs, 29864866 residues

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database: swiss-prot38

1:swissprot

Statistics: Mean 11.223; Variance 9.516; scale 1.179

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	16	100.0	68	1	ATP8_CERSI	2.99e+03
2	16	100.0	99	1	BMG_CAVPO	2.99e+03
3	16	100.0	145	1	DNBI_ADE02	2.99e+03
4	16	100.0	149	1	CYNS_SYNY3	2.99e+03
5	16	100.0	152	1	AVID_CHICK	2.99e+03
6	16	100.0	192	1	AMEY_HUMAN	2.99e+03
7	16	100.0	205	1	CYSR_SYNY3	2.99e+03
8	16	100.0	221	1	CD28_RABIT	2.99e+03
9	16	100.0	226	1	ATP6_BALMO	2.99e+03
10	16	100.0	241	1	CRTA_RHOCA	2.99e+03
11	16	100.0	242	1	CB8A_BOVIN	2.99e+03
12	16	100.0	242	1	DCOP_PENCH	2.99e+03
13	16	100.0	242	1	CRTW_AGRAU	2.99e+03
14	16	100.0	253	1	ADH_DROMM	2.99e+03
15	16	100.0	256	1	CHMO_YEAST	2.99e+03
16	16	100.0	258	1	DRG2_BACSU	2.99e+03
17	16	100.0	278	1	DCOP_PENCH	2.99e+03
18	16	100.0	281	1	ACAC9_MICCH	2.99e+03
19	16	100.0	297	1	APCH_RAT	2.99e+03
20	16	100.0	301	1	CRTB_AGRAU	2.99e+03
21	16	100.0	323	1	AMFR_HUMAN	2.99e+03
22	16	100.0	334	1	DCAM_HUMAN	2.99e+03
23	16	100.0	339	1	CYSM_ALCEU	2.99e+03

24	16	100.0	345	1	APDH_MOUSE	BETA-2-GLYCOPROTEIN I	2.99e+03
25	16	100.0	353	1	CUP5_GALME	PUPAL CUTICLE PROTEIN	2.99e+03
26	16	100.0	359	1	AROB_NEIGO	3-DEHYDROQUINATE SYNTH	2.99e+03
27	16	100.0	361	1	CIKE_HUMAN	VOLTAGE-GATED POTASSIU	2.99e+03
28	16	100.0	362	1	AROB_MYCTU	3-DEHYDROQUINATE SYNTH	2.99e+03
29	16	100.0	367	1	DCUP_MOUSE	UROPORPHYRINOGEN DECAR	2.99e+03
30	16	100.0	375	1	ACT_SCHPO	ACTIN.	2.99e+03
31	16	100.0	378	1	ACT_SCHDU	ACTIN.	2.99e+03
32	16	100.0	382	1	DCUP_MYCLE	UROPORPHYRINOGEN DECAR	2.99e+03
33	16	100.0	383	1	CYSL_SPTOL	CYSTINE SYNTHASE, CHL	2.99e+03
34	16	100.0	387	1	D4DR_MOUSE	D(4) DOPAMINE RECEPTOR	2.99e+03
35	16	100.0	392	1	CYB_VICFA	CYTOCHROME B.	2.99e+03
36	16	100.0	393	1	DCAM_HORCH	S-ADENOSYLMETHIONINE D	2.99e+03
37	16	100.0	488	1	DNAB_HELPY	REPLICATIVE DNA HELICA	2.99e+03
38	16	100.0	499	1	CHIB_SERMA	CHITINASE B PRECURSOR	2.99e+03
39	16	100.0	514	1	COX1_MOUSE	CYTOCHROME C OXIDASE P	2.99e+03
40	16	100.0	525	1	ACU8_NEUCR	ACETYL-COA HYDROLASE (	2.99e+03
41	16	100.0	573	1	COX1_MYCTU	PROBABLE CYTOCHROME C	2.99e+03
42	16	100.0	638	1	CIKE_RAT	VOLTAGE-GATED POTASSIU	2.99e+03
43	16	100.0	729	1	DD15_ARATH	PUTATIVE PRE-MRNA SPLI	2.99e+03
44	16	100.0	1132	1	DNBI_HSV6U	MAJOR DNA-BINDING PROT	2.99e+03
45	16	100.0	2032	1	CTOG_HUMAN	CH-TOG PROTEIN (COLONI	2.99e+03

## ALIGNMENTS

RESULT 1  
ID ATP8\_CERSI STANDARD; PRT; 68 AA.  
AC O03199;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DE ATP SYNTHASE PROTEIN 8 (EC 3.6.1.34) (A6L).  
GN WTATP8 OR ATP8  
OS Ceratotherium simum (White rhinoceros) (Square-lipped rhinoceros).  
OG Mitochondrion.  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Perissodactyla; Rhinocerotidae; Ceratotherium.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 97271844.  
RA XU X., ARNASON U.;  
RT "The complete mitochondrial DNA sequence of the white rhinoceros,  
RT Ceratotherium simum, and comparison with the mtDNA sequence of the  
RT Indian rhinoceros, Rhinoceros unicornis.";  
RL Mol. Phylogenet. Evol. 7:189-194(1997).  
CC -1- FUNCTION: THIS IS ONE OF THE CHAINS OF THE NONENZYMATIC COMPONENT  
CC (CF(0) SUBUNIT) OF THE MITOCHONDRIAL ATPASE COMPLEX.  
CC -1- SUBCELLULAR LOCATION: MEMBRANE-BOUND.  
CC -1- SIMILARITY: BELONGS TO THE ATPASE PROTEIN 8 FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC  
CC EMBL; Y0726; CAA69010.1; -  
CC PFAM; PF00895; ATP-synt\_8; 1.  
CC Hydrogen ion transport; CF(0); Mitochondrion; Transmembrane.  
CC TRANSMEM 8 24  
CC POTENTIAL.  
CC SEQUENCE 68 AA; 7892 MW; CC12DD9 CRC32;  
CC  
CC Query Match 100.0%; Score 16; DB 1; Length 68;  
CC Best Local Similarity 25.0%; Pred. No. 2.99e+03;  
CC Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 34 PSSPELXL 41

Qy 2 PXXXXXXL 9



Best Local Similarity 25.0%; Pred. No. 2.19e+03;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 201 PDDEIFRL 208

Qy 1 1  
2 PXXXXXXL 9

Search completed: Sat Apr 15 02:11:00 2000  
Job time : 18 secs.

04-Mar-1994  
A03805  
A93733  
Herisse, J.; Rigolet, M.; Dupont de Dinechin, S.; Gallibert, F.  
#journal Nucleic Acids Res. (1981) 9:4023-4042  
#title Nucleotide sequence of adenovirus 2 DNA fragment encoding for the carboxylic region of the fiber protein and the entire E4 region.  
#cross-references MUID:82059444  
#accession A03805  
#molecule\_type DNA  
#residues 1-294 #label HER  
#note this probable protein was assigned by correlating EM data and S1 digestion studies

GENETICS  
#map\_position 92.6-95.2  
CLASSIFICATION #superfamily adenovirus early E4 34K protein  
KEYWORDS early protein  
SUMMARY #length 294 #molecular-weight 34116 #checksum 8260

Query Match 100.0%; Score 16; DB 1; Length 294;  
Best Local Similarity 25.0%; Pred. No. 2.19e+03;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 66 PCSVGFTL 73  
Qy 2 PXXXXXXL 9

RESULT 14  
ENTRY NDBPT7  
TITLE exodeoxyribonuclease (EC 3.1.11.-) - phase T7  
ORGANISM #formal\_name phase T7  
DATE 13-Jun-1983 #sequence\_revision 26-Jul-1996 #text\_change  
REFERENCE S42316; A00780; S43608  
#authors Dunn, J.J.  
#submission submitted to the EMBL Data Library, October 1993  
#accession S42316  
#molecule\_type DNA  
#residues 1-300 #label DUN  
#cross-references EMBL:V01146; NID:g431187; PID:g431192  
REFERENCE A94615  
#authors Dunn, J.J.; Thompson, K.  
#submission submitted to the Nucleic Acid Sequence Database, September 1982  
#accession A00780  
#molecule\_type DNA  
#residues 'MSRDLVTIPRVNDIOGYIDSLERENSLKQMLEADYVAELEKLNGT', 4-300 #label DUF  
#note due to a frameshift error genes 5.9 and 6 were concatenated into a single reading frame

REFERENCE S42283  
#authors Dunn, J.J.; Studier, F.W.  
#journal J. Mol. Biol. (1983) 166:477-535  
#title Complete nucleotide sequence of bacteriophage T7 DNA and the locations of T7 genetic elements.  
#cross-references MUID:83241725  
#accession S43608  
#molecule\_type DNA  
#residues 'MSRDLVTIPRVNDIOGYIDSLERENSLKQMLEADYVAELEKLNGT', 4-300 #label DUW  
#cross-references EMBL:V01146  
#note due to a frameshift error genes 5.9 and 6 were concatenated into a single reading frame  
#note the authors did not translate the codon for residue 1

GENETICS  
#gene 6  
#map\_position 43.74-46.08  
FUNCTION #description 5' to 3' exonuclease specific for double-stranded DNA;

probably removes DNA-linked RNA primers  
essential for phage DNA replication; required for host DNA degradation and phage genetic recombination  
#superfamily phage T7 exodeoxyribonuclease  
#length 300 #molecular-weight 34502 #checksum 7095

Query Match 100.0%; Score 16; DB 1; Length 300;  
Best Local Similarity 25.0%; Pred. No. 2.19e+03;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 109 PVGYFEFL 116  
Qy 2 PXXXXXXL 9

RESULT 15  
ENTRY PBYG  
TITLE H+-transporting ATP synthase (EC 3.6.1.34) gamma chain -  
ORGANISM #formal\_name Synchocystis sp.  
#variety PCC 6803  
DATE 30-Sep-1991 #sequence\_revision 30-Sep-1991 #text\_change  
REFERENCE S08257; S17752; S74582; A34256; S14867  
#authors Werner, S.; Schumann, J.; Strotmann, H.  
#journal FEBS Lett. (1990) 261:204-208  
#title The primary structure of the gamma-subunit of the ATPase from Synchocystis 6803.  
#cross-references MUID:90169116  
#accession S08257  
#molecule\_type DNA  
#residues 1-314 #label WER  
#cross-references EMBL:Y07532; NID:g48007; PID:g48008  
REFERENCE S17745  
#authors Lill, H.; Nelson, N.  
#journal Plant Mol. Biol. (1991) 17:641-652  
#title The atp1 and atp2 operons of the cyanobacterium Synchocystis sp. PCC 6803  
#cross-references MUID:92003679  
#accession S17752  
#molecule\_type DNA  
#residues 1-314 #label LIL  
#cross-references EMBL:X58128; NID:g47506; PID:g47514  
REFERENCE S74322  
#authors Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakanura, Y.; Miyajima, N.; Hirose, M.; Sugita, M.; Sasamoto, S.; Kimura, T.; Hosouchi, T.; Matsuno, A.; Muraki, A.; Nakazaki, N.; Naruo, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda, M.; Tabata, S.  
#journal DNA Res. (1996) 3:109-136  
#title Sequence analysis of the genome of the unicellular cyanobacterium Synchocystis sp. PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions.  
#cross-references MUID:97061201  
#accession S74582  
#status nucleic acid sequence not shown; translation not shown  
#molecule\_type DNA  
#residues 1-314 #label KAN  
#cross-references EMBL:D90900; GB:AB001339; NID:g1651768; PID:d1017467; PID:g1651807  
#note the nucleotide sequence was submitted to the EMBL Data Library, June 1996

GENETICS  
#gene atpC  
CLASSIFICATION #superfamily H+-transporting ATP synthase gamma chain  
KEYWORDS ATP biosynthesis; hydrolase; membrane-associated complex  
SUMMARY #length 314 #molecular-weight 34605 #checksum 8352

Query Match 100.0%; Score 16; DB 1; Length 314;

```

COMMENT      This enzyme catalyzes the NADPH-dependent reduction of
              dihydrofolate to tetrahydrofolate.
GENETICS
#gene        SGD:DFR1
##cross-references SGD:S0005762; MIPS:YOR336W
#map_position 15R
CLASSIFICATION #superfamily type I dihydrofolate reductase; type I
               dihydrofolate reductase homology
KEYWORDS      NADP; oxidoreductase
FEATURE
8-132         #domain type I dihydrofolate reductase homology #label
              DFR\
34,38,68,74   #binding_site substrate (Glu, Phe, Arg) #status
              predicted
SUMMARY       #length 211 #molecular-weight 24261 #checksum 8536

Query Match      100.0%; Score 16; DB 1; Length 211;
Best Local Similarity 25.0%; Pred. No. 2.19e+03;
Matches          2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 172 PAOLKEFL 179
|
|
Qy 2 PXXXXXXL 9

RESULT 10
ENTRY   RDBEHS #type complete
TITLE   dihydrofolate reductase (EC 1.5.1.3) - saimirine herpesvirus
        1 (strain 488)
ORGANISM #formal_name saimirine herpesvirus 1
DATE      31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change
ACCESSION E34770
REFERENCE A34770
#authors   Biesinger, B.; Trimble, J.J.; Desrosiers, R.C.; Fleckenstein,
          B.
#journal   Virology (1990) 176:505-514
#title     The divergence between two oncogenic Herpesvirus saimiri
          strains in a genomic region related to the transforming
          phenotype.
#cross-references MUID:90266466
#accession E34770
#molecule_type DNA
#residues 1-213 #label BIE
##cross-references EMBL:M55264; NID:g331005; PID:g331010
CLASSIFICATION #superfamily type I dihydrofolate reductase; type I
               dihydrofolate reductase homology
KEYWORDS      methotrexate resistance; NADP; oxidoreductase; trimethoprim
          resistance
FEATURE
4-125         #domain type I dihydrofolate reductase homology #label
              DFR\
31,35,64,70   #binding_site substrate (Asp, Phe, Asn, Arg) #status
              predicted
SUMMARY       #length 213 #molecular-weight 24577 #checksum 1227

Query Match      100.0%; Score 16; DB 1; Length 213;
Best Local Similarity 25.0%; Pred. No. 2.19e+03;
Matches          2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 66 PLKDRINL 73
|
|
Qy 2 PXXXXXXL 9

RESULT 11
ENTRY   DNDPW #type complete
TITLE   repeat element protein - Campopletis sonorensis virus
ORGANISM #formal_name Campopletis sonorensis virus, Csv
#note     host Campopletis sonorensis (parasitic wasp); Heliothis
          virescens
DATE      31-Mar-1990 #sequence_revision 31-Mar-1990 #text_change
          21-Nov-1997

COMMENT      This enzyme catalyzes the NADPH-dependent reduction of
              dihydrofolate to tetrahydrofolate.
GENETICS
#gene        SGD:DFR1
##cross-references SGD:S0005762; MIPS:YOR336W
#map_position 15R
CLASSIFICATION #superfamily type I dihydrofolate reductase; type I
               dihydrofolate reductase homology
KEYWORDS      NADP; oxidoreductase
FEATURE
8-132         #domain type I dihydrofolate reductase homology #label
              DFR\
34,38,68,74   #binding_site substrate (Glu, Phe, Arg) #status
              predicted
SUMMARY       #length 211 #molecular-weight 24261 #checksum 8536

Query Match      100.0%; Score 16; DB 1; Length 211;
Best Local Similarity 25.0%; Pred. No. 2.19e+03;
Matches          2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 172 PAOLKEFL 179
|
|
Qy 2 PXXXXXXL 9

RESULT 10
ENTRY   RDBEHS #type complete
TITLE   dihydrofolate reductase (EC 1.5.1.3) - saimirine herpesvirus
        1 (strain 488)
ORGANISM #formal_name saimirine herpesvirus 1
DATE      31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change
ACCESSION E34770
REFERENCE A34770
#authors   Biesinger, B.; Trimble, J.J.; Desrosiers, R.C.; Fleckenstein,
          B.
#journal   Virology (1990) 176:505-514
#title     The divergence between two oncogenic Herpesvirus saimiri
          strains in a genomic region related to the transforming
          phenotype.
#cross-references MUID:90266466
#accession E34770
#molecule_type DNA
#residues 1-213 #label BIE
##cross-references EMBL:M55264; NID:g331005; PID:g331010
CLASSIFICATION #superfamily type I dihydrofolate reductase; type I
               dihydrofolate reductase homology
KEYWORDS      methotrexate resistance; NADP; oxidoreductase; trimethoprim
          resistance
FEATURE
4-125         #domain type I dihydrofolate reductase homology #label
              DFR\
31,35,64,70   #binding_site substrate (Asp, Phe, Asn, Arg) #status
              predicted
SUMMARY       #length 213 #molecular-weight 24577 #checksum 1227

Query Match      100.0%; Score 16; DB 1; Length 213;
Best Local Similarity 25.0%; Pred. No. 2.19e+03;
Matches          2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 66 PLKDRINL 73
|
|
Qy 2 PXXXXXXL 9

RESULT 11
ENTRY   DNDPW #type complete
TITLE   repeat element protein - Campopletis sonorensis virus
ORGANISM #formal_name Campopletis sonorensis virus, Csv
#note     host Campopletis sonorensis (parasitic wasp); Heliothis
          virescens
DATE      31-Mar-1990 #sequence_revision 31-Mar-1990 #text_change
          21-Nov-1997

ACCESSIONS    A31823
REFERENCE      Theilmann, D.A.; Summers, M.D.
#authors      Virology (1988) 167:329-341
#title        Identification and comparison of Campopletis sonorensis virus
              transcripts expressed from four genomic segments in the
              insect hosts Campopletis sonorensis and Heliothis virescens.
              #cross-references MUID:89073734
#accession    A31823
#molecule_type mRNA
#residues     1-235 #label THE
##cross-references GB:M23437; GB:M16988; NID:g323408; PID:g323409
COMMENT       The genome of this virus consists of at least 28 closed circular
              superhelical DNA segments; three of them contain homologous DNA
              sequences that code for one or several tandem-repeated element
              proteins.
CLASSIFICATION #superfamily parasitic wasp virus repeat element protein
FEATURE
57-235        #domain repeat element #label RPE
SUMMARY       #length 235 #molecular-weight 28044 #checksum 4181

Query Match      100.0%; Score 16; DB 1; Length 235;
Best Local Similarity 25.0%; Pred. No. 2.19e+03;
Matches          2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 105 PILGGIML 112
|
|
Qy 2 PXXXXXXL 9

RESULT 12
ENTRY   A26952 #type complete
TITLE   electron transfer flavoprotein beta chain homolog - Rhizobium
        meliloti
ALTERNATE_NAMES fixA protein
ORGANISM #formal_name Rhizobium meliloti
DATE      05-Oct-1988 #sequence_revision 12-Jul-1996 #text_change
ACCESSIONS    A26952
REFERENCE      Earl, C.D.; Ronson, C.W.; Ausubel, F.M.
#authors      J. Bacteriol. (1987) 169:1127-1136
#journal       Genetic and structural analysis of the Rhizobium meliloti
              fixA, fixB, fixC, and fixX genes.
#cross-references MUID:87137267
#accession    A26952
#molecule_type DNA
#residues     1-292 #label EAR
##cross-references GB:M15546; NID:g340664; PID:g551198
COMMENT       This protein is essential for symbiotic nitrogen fixation.
GENETICS
#gene         fixA
#note         operon contains fixA, fixB, fixC, and fixX genes
CLASSIFICATION #superfamily electron transfer flavoprotein beta chain
KEYWORDS       electron transfer; flavoprotein; nitrogen fixation
SUMMARY       #length 292 #molecular-weight 31146 #checksum 8924

Query Match      100.0%; Score 16; DB 1; Length 292;
Best Local Similarity 25.0%; Pred. No. 2.19e+03;
Matches          2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 63 PKMAEDAL 70
|
|
Qy 2 PXXXXXXL 9

RESULT 13
ENTRY   Q4ADC2 #type complete
TITLE   early E4 34K protein - human adenovirus 2
ORGANISM #formal_name Mastadenovirus h2 #common_name human adenovirus
          2
#note        host Homo sapiens (man)
DATE         02-Apr-1982 #sequence_revision 02-Apr-1982 #text_change

```

```
FEATURE
2-147
64 #domain globin homology #label GLB\
#binding site oxygen (His) (distal axial ligand) #status
predicted\
93 #binding_site heme iron (His) (proximal axial ligand)
#status predicted
SUMMARY #length 153 #molecular-weight 17016 #checksum 5051
Query Match 100.0%; Score 16; DB 1; Length 153;
Best Local Similarity 25.0%; Pred.No. 2.19e+03;
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
Db 22 PSHGQEV 29
|
|
QY 2 PXXXXXX 9
|
|
RESULT 7
ENTRY MYBAO #type complete
TITLE myoglobin - olive baboon (tentative sequence)
ORGANISM #formal_name Papio anubis; Papio hamadryas anubis
#common_name olive baboon
DATE 24-Apr-1984 #sequence_revision 27-Nov-1985 #text_change
14-Nov-1997
ACCESSIONS A90583; A02469
REFERENCE A90583
#authors Romero-Herrera, A.E.; Lehmann, H.
#journal Biochim. Biophys. Acta (1972) 278:465-481
#title The myoglobin of primates. III. Cercopithecidae (Old World
monkeys): Papio anubis (olive baboon) and Macaca
fascicularis (=irus, crab-eating monkey.
#cross-references MUID:73040318
#contents compositions of tryptic, peptic, and chymotryptic peptides
and sequences of residues 25, 110-102, 140, and 143-145
#accession A90583
#molecule_type protein
1-153 #label ROM
##residues the peptides were aligned by homology with the human
##note sequence
COMMENT This myoglobin was isolated from skeletal muscle.
CLASSIFICATION #superfamily globin; globin homology
KEYWORDS chromoprotein; heme; iron; muscle; oxygen carrier
FEATURE
2-147
64 #domain globin homology #label GLB\
#binding site oxygen (His) (distal axial ligand) #status
predicted\
93 #binding_site heme iron (His) (proximal axial ligand)
#status predicted
SUMMARY #length 153 #molecular-weight 17016 #checksum 5051
Query Match 100.0%; Score 16; DB 1; Length 153;
Best Local Similarity 25.0%; Pred.No. 2.19e+03;
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
Db 22 PSHGQEV 29
|
|
QY 2 PXXXXXX 9
|
|
RESULT 8
ENTRY fixr homolog - Agrobacterium tumefaciens (fragment)
TITLE #formal_name Agrobacterium tumefaciens
ORGANISM 19-Jul-1996 #sequence_revision 19-Jul-1996 #text_change
13-Nov-1998
ACCESSIONS I39709
REFERENCE I39709
#authors Matthysse, A.G.; White, S.; Lightfoot, R.
#journal J. Bacteriol. (1995) 177:1069-1075
#title Genes required for cellulose synthesis in Agrobacterium
tumefaciens.
#cross-references MUID:95164506
#accession I39709
```

```
##status preliminary; translated from GB/EMBL/DBJ
##molecule_type DNA
##residues 1-177 #label RES
##cross-references GB:L38609; NID:g710486; PID:g710487
CLASSIFICATION #superfamily ribitol dehydrogenase; short-chain alcohol
dehydrogenase homology
FEATURE
4-177
#domain short-chain alcohol dehydrogenase homology
(fragment) #label SADH
SUMMARY #length 177 #checksum 8599
Query Match 100.0%; Score 16; DB 2; Length 177;
Best Local Similarity 25.0%; Pred.No. 2.19e+03;
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
Db 74 PEGKLDAL 81
|
|
QY 2 PXXXXXX 9
|
|
RESULT 9
ENTRY RBYD #type complete
TITLE dihydrofolate reductase (EC 1.5.1.3) - yeast (Saccharomyces
cerevisiae)
ALTERNATE_NAMES protein O5231; protein YOR236w
ORGANISM #formal_name Saccharomyces cerevisiae
DATE 30-Sep-1990 #sequence_revision 30-Sep-1990 #text_change
12-Dec-1997
ACCESSIONS J02669; J0274; S06312; S67129
REFERENCE A91592
#authors Filing, M.E.; Kopf, J.; Richards, C.A.
#journal Gene (1988) 63:165-174
#title Nucleotide sequence of the dihydrofolate reductase gene of
Saccharomyces cerevisiae.
#cross-references MUID:88255864
#accession J02669
##molecule_type DNA
##residues 1-211 #label FLI
##cross-references GB:M18578; EMBL:M26667; NID:g171396; PID:g171397
REFERENCE A91593
#authors Barclay, B.J.; Huang, T.; Nagel, M.G.; Misener, V.L.; Game,
J.C.; Wahl, G.M.
#journal Gene (1988) 63:175-185
#title Mapping and sequencing of the dihydrofolate reductase gene
(DHFR) of Saccharomyces cerevisiae.
#cross-references MUID:88255865
#accession J0274
##molecule_type DNA
##residues 1-211 #label BAR
##cross-references EMBL:M26668; NID:g295603; PID:g295604
REFERENCE S06312
#authors Lagosky, P.A.; Taylor, G.R.; Haynes, R.H.
#journal Nucleic Acids Res. (1987) 15:10355-10371
#title Molecular characterization of the Saccharomyces cerevisiae
dihydrofolate reductase gene (DHFR).
#cross-references MUID:88096572
#accession S06312
##molecule_type DNA
##residues 1-211 #label LAG
##cross-references EMBL:Y00887
##note the authors translated the codon GTA for residue 27 as
Leu; the sequence shown follows the authors'
translation
REFERENCE S67104
#authors Boyer, J.; Fairhead, C.; Gaillon, L.; Gallisson, F.; Michaux,
G.; Thierry, A.; Dujon, B.
#submission submitted to the Protein Sequence Database, July 1996
#accession S67129
#molecule_type DNA
##residues 1-211 #label BOY
##cross-references EMBL:Z75144; NID:g1420540; PID:e252096; PID:g1420541;
MIPS:YOR236w
##experimental_source strain S288C
```

DATE 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change  
08-Sep-1997  
ACCESSIONS S17343  
REFERENCE S17343  
#authors Corrochano, L.M.; Avalos, J.  
#description submitted to the EMBL Data Library, August 1991  
#description Cloning a segment of the 3-hydroxy-3-methylglutaryl coenzyme A  
reductase gene of *Phycomyces blakesleeanus* and *Gibberella*  
*fujikuroi*.  
#accession S17343  
#status preliminary  
#molecule\_type DNA  
#residues 1-106 #label COR  
#cross-references EMBL:X58370; NID:g2739; PID:g2740  
GENETICS  
#gene hmgA  
CLASSIFICATION #superfamily hydroxymethylglutaryl-CoA reductase (NADPH)  
KEYWORDS NADP; oxidoreductase  
SUMMARY #length 106 #molecular-weight 11019 #checksum 6079  
Query Match 100.0%; Score 16; DB 2; Length 106;  
Best Local Similarity 25.0%; Pred. No. 2.19e+03;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
Db 39 PCVAFETL 46  
| |  
QY 2 PXXXXXXL 9  
RESULT 3 QOEC25 #type complete  
ENTRY hypothetical 12.5K protein (trbB-trbF intergenic region) -  
TITLE Escherichia coli plasmid F  
ORGANISM #formal\_name Escherichia coli  
DATE 30-Jun-1990 #sequence\_revision 30-Jun-1990 #text\_change  
ACCESSIONS F32238  
REFERENCE A32238  
#authors Wu, J.H.; Ippen-Ihler, K.  
#journal J. Bacteriol. (1989) 171:213-221  
#title Nucleotide sequence of traQ and adjacent loci in the  
Escherichia coli K-12 F-plasmid transfer operon.  
#cross-references MUID:89123020  
#accession F32238  
#molecule\_type DNA  
#residues 1-113 #label WU1  
#experimental\_source strain K12  
GENETICS  
#genome plasmid  
SUMMARY #length 113 #molecular-weight 12587 #checksum 1549  
Query Match 100.0%; Score 16; DB 1; Length 113;  
Best Local Similarity 25.0%; Pred. No. 2.19e+03;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
Db 59 PVEIVWSL 66  
| |  
QY 2 PXXXXXXL 9  
RESULT 4 QOPSHT #type complete  
ENTRY hypothetical protein merr - Pseudomonas aeruginosa transposon  
TITLE Tn501  
ORGANISM #formal\_name Pseudomonas aeruginosa  
DATE 04-Dec-1986 #sequence\_revision 04-Dec-1986 #text\_change  
ACCESSIONS A04457  
REFERENCE A03556  
#authors Misra, T.K.; Brown, N.L.; Fritzinger, D.C.; Pridmore, R.D.;  
Barnes, W.M.; Haberstroh, L.; Silver, S.  
#journal Proc. Natl. Acad. Sci. U.S.A. (1984) 81:5975-5979  
#title Mercuric ion-resistance operons of plasmid R100 and

transposon Tn501: the beginning of the operon including the  
regulatory region and the first two structural genes.  
#cross-references MUID:85014891  
#accession A04457  
#molecule\_type DNA  
#residues 1-116 #label MIS  
GENETICS  
#gene merr  
CLASSIFICATION #superfamily merr protein  
KEYWORDS transmembrane protein  
SUMMARY #length 116 #molecular-weight 12498 #checksum 732  
Query Match 100.0%; Score 16; DB 1; Length 116;  
Best Local Similarity 25.0%; Pred. No. 2.19e+03;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
Db 4 PKTGRGAL 11  
| |  
QY 2 PXXXXXXL 9  
RESULT 5 WZBE32 #type complete  
ENTRY gene 32 protein - human herpesvirus 3  
TITLE #formal\_name human herpesvirus 3, varicella-zoster virus  
ORGANISM 30-Sep-1988 #sequence\_revision 30-Sep-1988 #text\_change  
DATE 14-Nov-1997  
ACCESSIONS F27214  
REFERENCE A27345  
#authors Davison, A.J.; Scott, J.E.  
#journal J. Gen. Virol. (1986) 67:1759-1816  
#title The complete DNA sequence of varicella-zoster virus.  
#cross-references MUID:86306657  
#accession F27214  
#molecule\_type DNA  
#residues 1-143 #label DAV  
#cross-references EMBL:X04370; NID:g59989; PID:g60021  
GENETICS  
#gene 32  
CLASSIFICATION #superfamily varicella-zoster virus gene 32 protein  
SUMMARY #length 143 #molecular-weight 15981 #checksum 7527  
Query Match 100.0%; Score 16; DB 1; Length 143;  
Best Local Similarity 25.0%; Pred. No. 2.19e+03;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
Db 36 PAIEDRL 43  
| |  
QY 2 PXXXXXXL 9  
RESULT 6 MYMQRG #type complete  
ENTRY myoglobin - red guenon (tentative sequence)  
TITLE #formal\_name Erythrocybus patas #common\_name red guenon,  
ORGANISM hussar  
DATE 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change  
ACCESSIONS A90633; A02469  
REFERENCE A90633  
#authors Dene, H.; Sazy, J.; Romero-Herrera, A.E.  
#journal Biochim. Biophys. Acta (1980) 625:133-145  
#title The myoglobin of primates. X.  
#cross-references MUID:81021734  
#contents compositions of tryptic and peptic peptides and sequences of  
residues 17-23, 103-105, 110, and 140-142  
#accession A90633  
#molecule\_type protein  
#residues 1-153 #label DEN  
#note the peptides were aligned by homology with the human  
sequence  
COMMENT This myoglobin was isolated from skeletal muscle.  
CLASSIFICATION #superfamily globin; globin homology  
KEYWORDS chromoprotein; heme; iron; muscle; oxygen carrier

\*\*\*\*\*  
WPSRFL  
\*\*\*\*\*  
(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Sat Apr 15 02:10:42 2000; MasPar time 3.05 seconds  
Tabular output not generated. 118.324 Million cell updates/sec

Title: >US-08-452-843-29  
Description: (1-9) from US08452843.pap  
Sequence: 1 XPXXXXXXL 9

Scoring table: PAM 150  
Gap 15

Searched: 122810 seqs, 40068593 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: p1r62  
1:p1r1 2:p1r2 3:p1r3 4:p1r4

Statistics: Mean 10.355; Variance 8.040; scale 1.288

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	16	100.0	25	2	cytochrome P450LPGA O	2.19e+03
2	16	100.0	106	2	hydroxymethylglutaryl	2.19e+03
3	16	100.0	113	1	hypothetical 12.5K pr	2.19e+03
4	16	100.0	116	1	hypothetical protein	2.19e+03
5	16	100.0	143	1	gene 32 protein - hum	2.19e+03
6	16	100.0	153	1	myoglobin - red gueno	2.19e+03
7	16	100.0	153	1	myoglobin - olive bab	2.19e+03
8	16	100.0	177	2	fixR homolog - Agrob	2.19e+03
9	16	100.0	211	1	dihydrofolate reducta	2.19e+03
10	16	100.0	213	1	rod shape-determining	2.19e+03
11	16	100.0	235	1	repeat element protei	2.19e+03
12	16	100.0	292	1	early E4 34K protein	2.19e+03
13	16	100.0	294	1	exodeoxyribonuclease	2.19e+03
14	16	100.0	300	1	H+-transporting ATP s	2.19e+03
15	16	100.0	314	1	tryptophan--trNA liga	2.19e+03
16	16	100.0	328	1	fructose-bisphosphata	2.19e+03
17	16	100.0	341	1	rod shape-determining	2.19e+03
18	16	100.0	370	1	alcohol dehydrogenase	2.19e+03
19	16	100.0	375	1	monocyte surface glyc	2.19e+03
20	16	100.0	375	1	coat protein VP1 - mo	2.19e+03
21	16	100.0	384	1	immediate-early prote	2.19e+03
22	16	100.0	391	1	polyketide synthase (	2.19e+03
23	16	100.0	411	2		

24	16	100.0	412	2	B40634	erythromycin monooxyg	2.19e+03
25	16	100.0	417	1	VGBEIB	glycoprotein D precu	2.19e+03
26	16	100.0	433	1	DEECS	homoserine dehydrogen	2.19e+03
27	16	100.0	438	1	XXNSN	phosphatidylcholine--	2.19e+03
28	16	100.0	458	1	WMSRL	biliary glycoprotein	2.19e+03
29	16	100.0	467	1	HLMSF3	poliovirus receptor h	2.19e+03
30	16	100.0	508	2	A36304	cytochrome P450 4A8 -	2.19e+03
31	16	100.0	518	1	SYECEC	glutamate--cysteine 1	2.19e+03
32	16	100.0	526	2	JC4533	cytochrome P450 4F5 p	2.19e+03
33	16	100.0	537	2	JC4534	cytochrome P450 4F6 p	2.19e+03
34	16	100.0	585	1	SHADH5	peripentonal hexon-as	2.19e+03
35	16	100.0	604	1	QXXLSM	NADH dehydrogenase (u	2.19e+03
36	16	100.0	606	1	QXBOSM	NADH dehydrogenase (u	2.19e+03
37	16	100.0	745	1	H64653	copper-transporting A	2.19e+03
38	16	100.0	789	1	QXBYS2	gene coxI intron 2 pr	2.19e+03
39	16	100.0	808	1	QPKEX	glucose dehydrogenase	2.19e+03
40	16	100.0	834	1	QXBYS1	gene coxI intron 1 pr	2.19e+03
41	16	100.0	901	1	WNVNTN	104K glycoprotein - T	2.19e+03
42	16	100.0	1015	1	JS0628	formate dehydrogenase	2.19e+03
43	16	100.0	1016	1	S40838	formate dehydrogenase	2.19e+03
44	16	100.0	2314	1	A46151	protein-tyrosine-phos	2.19e+03
45	16	100.0	2351	1	EZHU	coagulation factor VI	2.19e+03

ALIGNMENTS

RESULT 1  
ENTRY P00048 #type fragment  
TITLE cytochrome P450LPGA omega 2 - rabbit (fragment)  
CONTAINS oxidoreductase (EC 1.-.-.-)  
ORGANISM #formal\_name Oryctolagus cuniculus #common\_name domestic rabbit  
DATE 31-Dec-1990 #sequence\_revision 31-Dec-1990 #text\_change 05-Mar-1999

ACCESSIONS P00048  
REFERENCE P00047  
#authors Kikuta, Y.; Kusunose, E.; Okumoto, T.; Kubota, I.; Kusunose, M.  
#journal J. Biochem. (1990) 107:280-286  
#title Purification and characterization of two forms of cytochrome P-450 with omega-hydroxylase activities toward prostaglandin A and fatty acids from rabbit liver microsomes.

#cross-references MUID:90299866  
#accession P00048  
#molecule\_type protein  
#residues 1-25 #label KIK  
#experimental\_source liver  
#COMMENT This enzyme catalyzes the omega-hydroxylation of prostaglandin A1 and A2, as well as the omega- and (omega-1)-hydroxylation of fatty acid.

GENETICS CYP4A  
CLASSIFICATION #superfamily human cytochrome P450 CYP4B1; cytochrome P450 homology  
KEYWORDS electron transfer; heme; monooxygenase; oxidoreductase; transmembrane protein  
SUMMARY #length 25 #checksum 4616

Query Match 100.0%; Score 16; DB 2; Length 25;  
Best Local Similarity 25.0%; Pred. No. 2.19e+03;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 8 PGSPSGFL 15  
Qy 2 PXXXXXXL 9

RESULT 2  
ENTRY S17343 #type complete  
TITLE hydroxymethylglutaryl-CoA reductase (NADPH) (EC 1.1.1.34) - fungus (Gibberella fujikuroi)  
ORGANISM #formal\_name Gibberella fujikuroi, Fusarium moniliforme

THIS PAGE BLANK (USPTO)

QY 2 PXXXXXXL 9

RESULT 13  
ID R23769 standard; Protein; 134 AA.  
AC R23769;  
DT 27-OCT-1992 (first entry)  
DE Recombinant light chain variable domain (3).  
KW Complementarity determining region; light chain variable domain;  
KW antigen binding site; ligand; framework region; cancer; transplant.  
OS Synthetic.  
FH Key  
FT region Location/Qualifiers  
FT 23..35  
FT /label= CDR(d)  
FT 51..63  
FT /label= CDR(d)  
FT 98..104  
FT /label= CDR(e)  
FN W09206193-A.  
PD 16-APR-1992.  
PF 04-OCT-1991; G01726.  
PR 05-OCT-1990; GB-021679.  
PA (Gorman SD, Routledge EG, Waldmann H;  
WPI; 92-150879/18.  
DR Ligands and antibodies with binding affinity for CD3 antigen -  
PT for treatment of immunosuppression e.g. in graft rejection, and  
PT cancer, esp. lymphoid malignancies  
PS Claim 7; Page 31; 49pp; English.  
CC The sequence given is a recombinant human light chain variable  
CC domain ligand containing the complementarity determining region  
CC (CDR) given in R23736 and R23737. CDR's are found in the variable  
CC domains of light and heavy chains which form the antigen binding site,  
CC and act as connectors between the four framework regions.  
CC It has been noted that there seem to be no characteristic features  
CC which distinguish human from mouse or rat CDR's and they are  
CC therefore immunologically identical. This ligand has binding affinity  
CC for the human CD3 antigen and due to the lack of immunological  
CC response caused by the synthetic CDR's the ligand can be considered to  
CC be humanised. This ligand can be used to manufacture medicaments  
CC for use in immunosuppression esp. in patients with cancer or transplant  
CC recipients.  
SQ Sequence 134 AA;

Query Match 100.0%; Score 16; DB 1; Length 134;  
Best Local Similarity 25.0%; Pred. No. 1.71e+03;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 45 PTTVIFTL 52

QY 2 PXXXXXXL 9

RESULT 14  
ID P91916 standard; protein; 154 AA.  
AC P91916;  
DT 14-MAY-1990 (first entry)  
DE Derived sequence of the coding region of a cDNA clone encoding  
DE murine interleukin-7 (mIL-7)  
KW Murine interleukin-7 (mIL-7); murine stromal cell culture;  
KW lymphopoietic response; hematopoietic response.  
OS Mouse.  
FH Key  
FT peptide Location/Qualifiers  
FT 1..75  
FT /note= "putative signal peptide"  
FT 76..154  
FN protein  
PD EP-314415-A.  
PD 03-MAY-1989.  
PF 24-OCT-1988; 309977  
PR 26-OCT-1987; US-113566.  
PA (IMMU-) Immunex Corp.  
PI Namen AE, Goodwin RG, Lupton SD, Mochizuki DY, Price VL, Deeley MG;  
DR WPI; 89-131937/18.

DR N-PSDB; N90781.  
PT Mammalian interleukin-7, analogues and sub-units  
PT - used for modulating or augmenting immune,  
PT lymphopoietic and/or hematopoietic response in mammals  
PS Figure 3; 32pp; English.  
CC The coding sequence was identified by direct expression cloning using a  
CC mammalian expression vector. A putative factor observed in murine stromal  
CC cell cultures was designated IL-7. A novel cell line was established by  
CC transformation of stromal cells and this cell line provided specific  
CC IL-7 messenger RNA for expression, cloning and protein for purification  
CC and sequencing. A purified IL-7 protein compsn. pref. has a specific  
CC activity of more than 1x10(4) IL-7 units/microgram. It can be used for  
CC modulating or augmenting immune, lymphopoietic and/or hematopoietic  
CC responses in mammals.  
SQ Sequence 154 AA;

Query Match 100.0%; Score 16; DB 1; Length 154;  
Best Local Similarity 25.0%; Pred. No. 1.71e+03;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 13 PPLILVLL 20

QY 2 PXXXXXXL 9

RESULT 15  
ID P70262 standard; protein; 170 AA.  
AC P70262;  
DT 27-FEB-1991 (first entry)  
DE Beta-glucuronidase.  
KW Plasmid pBG1; beta-glucuronidase.  
PN EP-234075-A.  
PD 02-SEP-1987.  
PF 24-FEB-1986; 301305.  
PR 24-FEB-1986; EP-301305.  
PA (REPL-) Repligen Corporation.  
PI Anilionis A, Palmer JL;  
DR WPI; 87-243705/35.  
DR N-PSDB; N70396.  
PT New DNA sequences for protein prodn. esp. beta-glucuronidase - obtd.  
PT using beta-glucuronidase gene promoter DNA for high level expression  
PT in Escherichia coli.  
PS Claim 7; page 15; 19pp; English.  
CC The beta-glucuronidase coding sequence is carried in plasmid pBG1.  
CC It is expressed in high levels, eg over 50% of the total cellular  
CC protein of the host may comprise this single protein prod  
CC The coding sequence may be fused to a DNA sequence coding for a  
CC different protein.  
SQ Sequence 170 AA;

Query Match 100.0%; Score 16; DB 1; Length 170;  
Best Local Similarity 25.0%; Pred. No. 1.71e+03;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 8 PTREIKKL 15

QY 2 PXXXXXXL 9

Search completed: Sat Apr 15 02:10:24 2000  
Job time : 38 secs.



PR 12-APR-1988; GB-008524.  
PA (BRBI-) BRITISH BIO-TECHN L.  
PI Edwards RM;  
DR WPI: 89-311767/43.  
DR N-PSDB; N91647.  
PT Synthetic gene encoding human interleukin-5 - has restriction  
PT sites at frequent intervals to facilitate manipulation  
PS Disclosure; Fig 3a; 21pp; English.  
CC N91647 has restriction sites for HindIII, BspMI, NcoI, SpeI, BspMI,  
CC ApaLI, XmnI, ClaI, BalI, PstI, DraIII, BamHI and EcoRI. IL5 acts as  
CC a B-cell growth and differentiation factor.  
SQ Sequence 113 AA;

Query Match 100.0%; Score 16; DB 1; Length 113;  
Best Local Similarity 25.0%; Pred. No. 1.71e+03;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 34 PVKHNQQL 41  
|  
QY 2 PXXXXXXL 9

RESULT 10  
ID R24108 standard; Protein: 126 AA.  
AC R24108;  
DT 25-NOV-1992 (first entry)  
DE Humanised anti-Tac antibody light chain.  
KW Immunoglobulin; T cell related diseases; leukaemia; autoimmune;  
KW IL-2 receptor; recombinant; diagnosis; therapy.  
OS Synthetic.  
FH Key Location/Qualifiers  
FT peptide 1..20  
FT peptide /note= "signal peptide"  
FT peptide 20..126  
FT peptide /note= "mature peptide"  
PN DD-296964-A.  
PD 19-DEC-1991.  
PF 17-JAN-1990; 337159.  
PR 17-JAN-1990; DD-337159.  
PA (PROT-) PROTEIN DESIGN LABS INC.  
PI Queen CL, Sellick HE;  
DR WPI: 92-167794/21.  
DR N-PSDB; Q24791.

PT New humanised antibody specific for interleukin-2 receptor - with  
PT complementarity determin. regions and framework from different  
PT immunoglobulin(s), is non immunogenic and used to treat T-cell  
PT mediated disorders  
PS Disclosure; Fig 4; 21pp; German.  
CC The sequence is that of the humanised anti-Tac antibody light chain  
CC which is used in the production of a human-type immunoglobulin (Ig)  
CC that reacts specifically with p55-Tac protein and/or inhibits binding  
CC of human interleukin-2 (IL-2) to its specific receptor. The three  
CC complementarity determining regions and amino acids 48, 60 and 63  
CC of human antibody Eu have been replaced with the corresponding amino  
CC acids in the anti-Tac heavy chain antibody. This produces a humanised  
CC antibody which has the same affinity as anti-Tac for IL-2 receptors.  
CC The Ig may be used to treat humans with T-cell related diseases such  
CC as transplant rejection, T cell leukaemia, or autoimmune diseases such  
CC as diabetes or multiple sclerosis. See also R24103-R24107.  
SQ Sequence 126 AA;

Query Match 100.0%; Score 16; DB 1; Length 126;  
Best Local Similarity 25.0%; Pred. No. 1.71e+03;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 59 PGKAPKLL 66  
|  
QY 2 PXXXXXXL 9

RESULT 11  
ID R22481 standard; Protein: 130 AA.  
AC R22481;

DT 22-SEP-1992 (first entry)  
DE Neurotrophic factor 4 activity variants.  
KW NT-4; NT-3; BDNF; NGF; mutagenesis; substitution.  
OS Homo sapiens  
FH Key Location/Qualifiers  
FT misc\_difference 116..116  
FT /note= "GLU, ASN, GLN, TYR, SER, THR"  
PN WO9205254-A.  
PD 02-APR-1992.  
PF 24-SEP-1991; U06950.  
PR 25-SEP-1990; US-587707.  
PR 31-JAN-1991; US-648482.  
PA (GETH) GENENTECH INC.  
PI Rosenthal A;  
DR WPI: 92-132123/16.  
DR Neurotrophic factor-4 - useful for treating neurodegenerative  
DR diseases e.g. Alzheimer's and Parkinson's diseases, nerve cells  
DR damaged by e.g. diabetes  
PS Disclosure; Seq 68-73; 84pp; English.  
CC The sequence shows a portion of the amino acid sequence of human  
CC neurotrophic factor-4 (NT-4), (full sequence R22465). Position 116  
CC is a point at which substitution mutation causes a marked  
CC differentiation in the activity of the trophic element. Either Glu,  
CC Asn, Gln, Tyr, Ser or Thr may be included at this point. The sites  
CC of greatest interest for substitutional mutagenesis include sites  
CC where the amino acids found in BDNF, NGF, NT-3, and NT-4 are  
CC substantially different in terms of side chain bulk, charge, or  
CC hydrophobicity, but where there is also a high degree of homology at  
CC the selected site within various animal analogues of NGF, NT-3 and  
CC BDNF.  
SQ Sequence 130 AA;

Query Match 100.0%; Score 16; DB 1; Length 130;  
Best Local Similarity 25.0%; Pred. No. 1.71e+03;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 7 PASRRGEL 14  
|  
QY 2 PXXXXXXL 9

RESULT 12  
ID R22821 standard; Protein: 134 AA.  
AC R22821;  
DT 04-SEP-1992 (first entry)  
DE R99, K113, E122, M127, E132, S134 hIL-31.  
KW Human; interleukin-3; bone marrow transplant; graft; platelet;  
KW hIL-31; derivative.  
OS Homo sapiens.  
PN J04063595-A.  
PD 28-FEB-1992.  
PF 19-OCT-1990; 279108.  
PR 03-APR-1990; JP-087468.  
PR 19-OCT-1990; JP-279108.  
PA (KIRI) KIRIN BREWERY KK.  
DR WPI: 92-120155/15.  
DR N-PSDB; Q22513.  
PT Human interleukin 3 deriv. and its prepn. - for supplementing  
PT bone marrow transplantation and increasing platelet count  
PS Disclosure; Fig 6; 21pp; Japanese.  
CC This sequence codes for a derivative of human IL-3 having amino acids  
CC His, Thr, Ala, Thr, Ala and Phe at positions 99, 113, 122, 127, 132  
CC and 134, respectively, substituted by Arg, Lys, Glu, Met, Glu and Ser.  
CC The derivative has a higher activity than native IL-3.  
CC See R22813-4 and Q22503-Q22510.  
SQ Sequence 134 AA;

Query Match 100.0%; Score 16; DB 1; Length 134;  
Best Local Similarity 25.0%; Pred. No. 1.71e+03;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 3 PMTQTSL 10  
|

PT Poly:peptide with granulocyte colony stimulating factor activity  
 PT - obtd. by recombinant DNA procedures for treating haematopoietic  
 PT disorders  
 PS Example; p11; 79pp; English.  
 CC The examples describe procedures for the designing of probes for  
 CC hpG-CSF cDNA and genomic clones, both of which are claimed.  
 CC Specifically claimed are DNA sequences encoding for (Ala 1)hpG-CSF;  
 CC (Ser 36, 42, 64 and 74)hpG-CSF and the corresponding Met-1 cpds.  
 CC The novelty is that hpG-CSF is the prod. of procaryotic or  
 CC eucaryotic expression of an exogenous DNA sequence. The construction  
 CC of hpG-CSF expression vectors is also described in the examples.  
 SQ Sequence 22 AA;

Query Match 100.0%; Score 16; DB 1; Length 22;

Best Local Similarity 37.5%; Pred. No. 1.71e+03; Mismatches 5; Indels 0; Gaps 0;

Db 2 PLGPASXL 9  
 |  
 |  
 QY 2 PXXXXXXL 9

## RESULT 6

ID P81221 standard; protein; 71 AA.

AC P81221;

DT 07-DEC-1990 (first entry)

DE Methionine-insulin-like growth factor 1.

KW Methionine-insulin-like growth factor I;

KW two-cistronic Met-IGF-I expression vector.

OS Synthetic.

PN EP-264074-A.

PD 20-APR-1988.

PF 08-OCT-1987; 114733.

PR 09-OCT-1986; JP-240702.

PA (FUJI ) FUJISAWA PHARM KK.

PI Ikuro U, Mineo N, Yoshimasa S, Yoshinori I, Tadashi K;

DR WPI; 88-106856/16.

DR N-PSDB; N81582.

PT Prepn. of methionine-insulin-like growth factor I - comprises use

PT of recombinant DNA expression vector as transformant in

PT Escherichia coli.

PS Disclosure; p; English.

CC A two-cistronic vector functional and replicatable in E. coli,

CC which essentially contains DNA encoding IGF-1 and a protective peptide

CC capable of preventing the cellular proteases from decomposing IGF-1

CC was constructed.

CC See also N81563-82.

SQ Sequence 71 AA;

Query Match 100.0%; Score 16; DB 1; Length 71;

Best Local Similarity 25.0%; Pred. No. 1.71e+03; Mismatches 6; Indels 0; Gaps 0;

Db 51 PSCDLRRL 58  
 |  
 |  
 QY 2 PXXXXXXL 9

## RESULT 7

ID P91090 standard; Protein; 83 AA.

AC P91090;

DT 13-MAR-1992 (first entry)

DE Sequence of viper venom polypeptide called "Bitistatin 1".

KW Platelet aggregation inhibitor; antithrombotic agent;

KW myocardial infarction.

OS Viper.

PN EP-338634-A.

PD 25-OCT-1989.

PF 17-APR-1989; 200967.

PR 22-APR-1988; US-184653.

PR 22-APR-1988; US-184649.

PR 01-FEB-1989; US-303757.

PA (MERI ) MERCK & CO INC.

PI Friedman PA, Polokoff MA, Gould RJ, Bencen GH, Jacobs JW,  
 PI Garsky VM, Gan ZR;  
 DR WPI; 89-311082/43.  
 PT: Viper venom polypeptide cpds. - useful in inhibiting platelet  
 PT aggregation where strong antithrombotic activity of short  
 PT duration is needed  
 PS Claim 5; Page 22; 33pp; English.  
 CC The polypeptides of the invention have been purified from the venom  
 CC of various vipers, e.g. Trimeresurus gramineus, E. carinatus,  
 CC Agkistrodon piscivorus, Bitis arietans and Eristocophis macmahonii.  
 CC The polypeptides can be used to prevent platelet thrombosis,  
 CC thromboembolism and reocclusion.  
 SQ Sequence 83 AA;

Query Match 100.0%; Score 16; DB 1; Length 83;

Best Local Similarity 25.0%; Pred. No. 1.71e+03; Mismatches 6; Indels 0; Gaps 0;

Db 2 PPVCGNEL 9  
 |  
 |  
 QY 2 PXXXXXXL 9

## RESULT 8

ID P83143 standard; protein; 108 AA.

AC P83143;

DT 20-NOV-1990 (first entry)

DE Sequence of novel pre-S1 region from hepatitis B virus (HBV)

DE adw subtype

KW Vaccine; antigen; immunogen; probe; hybridisation;

KW Immunooassay; diagnosis.

OS Hepatitis B virus adw subtype.

PN EP-278940-A.

PD 17-AUG-1988.

PF 25-JAN-1988; 870008.

PR 30-JAN-1987; US-009325.

PA (SMIK) Smith Kline-Rit.

PI Cabezon T, De Wilde M, Harford N;

DR WPI; 88-229751/33.

DR N-PSDB; n81106.

PT DNA encoding hepatitis B virus antigens and hybrids contg. them -

PT used for expression in yeast to obtain vaccines and bivalent

PT vaccines

PS Claim 21; Pages 55; 101pp; English.

CC The DNA encoding HBV Pre-S2 (n81105) and Pre-S1 (n81106) regions was

CC isolated from plasmid pRIT10616 (ATCC 38131). HBV Pre-S1 protein (P80505)

CC and Pre-S2 protein (P80505) coding regions are claimed. The DNA sequences

CC are expressed in yeasts to obtain proteins or hybrid polypeptides which

CC are useful in the prepn. of vaccines and bivalent vaccines eg to HBV and

CC malaria. The DNA fragments and HBVag produced by it can also be used as a

CC probe for detection of HBV in biological samples by DNA hybridisation and

CC various immunoassays.

SQ Sequence 108 AA;

Query Match 100.0%; Score 16; DB 1; Length 108;

Best Local Similarity 25.0%; Pred. No. 1.71e+03; Mismatches 6; Indels 0; Gaps 0;

Db 94 PTFISPPPL 101  
 |  
 |  
 QY 2 PXXXXXXL 9

## RESULT 9

ID P93152 standard; Protein; 113 AA.

AC P93152;

DT 15-MAR-1992 (first entry)

DE Sequence of human interleukin-5 (IL-5).

KW B-cell growth factor; lymphokine; B-cell stimulating factor 2.

OS Homo sapiens.

PN GB2217328-A.

PD 25-OCT-1989.

PF 12-APR-1988; 008524.

RESULT 2  
ID R10604 standard; Protein; 17 AA.  
AC R10604; 1991 (first entry)  
DT 18-APR-1991 (first entry)  
DE Peptide with motilin-like activity (G).  
KW Motilin; activity; gastrointestinal disorder; drug.  
OS Synthetic.  
PN J02311495-A.  
PD 27-DEC-1990. 128911.  
PF 24-MAY-1989; 128911.  
PR 24-MAY-1989; JP-128911.  
PA (SANWA) SANWA KAGAKU KENKYUSHO.  
DR WPI: 91-047299/07.  
PT Polypeptide(s) with motilin-like activity - used as active  
component of drug for treating gastrointestinal disorder  
PS Disclosure; Page 4; 7pp; Japanese.  
CC Compared with motilin, the peptide chain is considerably shorter.  
CC Chemical synthesis is easy and cheap. The peptide is used  
as active component in a drug for treating gastrointestinal disorders.  
CC See also R10598-R10611.  
SQ Sequence 17 AA;

Query Match 100.0%; Score 16; DB 1; Length 17;  
Best Local Similarity 25.0%; Pred. No. 1.71e+03;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 3 PIFRYGEL 10  
QY 2 PXXXXXXL 9

RESULT 3  
ID P70726 standard; protein; 17 AA.  
AC P70726;  
DT 26-APR-1991 (first entry)  
DE Sequence of N-terminal of human granulocyte colony stimulating  
factor (hpg-CSF).  
KW Haematopoietic disorders; therapy; aplastic anaemia;  
bone marrow transplant; burn wounds; leukaemia.  
OS Homo sapiens.  
FH Key Location/Qualifiers  
FT misc\_difference 12  
PN W08701132-A.  
PD 26-FEB-1987.  
PF 22-AUG-1986; U01708.  
PR 23-AUG-1985; US-768954.  
PR 23-AUG-1985; US-768959.  
PR 03-MAR-1986; US-835548.  
PA (KIRI-) KIRIN-AMGEN INC.  
PA (KIRI) KIRIN-AMGEN INC.  
PI Souza LM;  
DR WPI: 87-064855/09.  
PT Poly-peptide with granulocyte colony stimulating factor activity  
- obt'd. by recombinant DNA procedures for treating haematopoietic  
disorders  
PT disorders  
PS Example; p10; 79pp; English.  
CC The examples describe procedures for the designing of probes for  
hpg-CSF cDNA and genomic clones, both of which are claimed.  
CC Specifically claimed are DNA sequences encoding for (Ala 1)hpg-CSF;  
(Ser 36, 42, 64 and 74)hpg-CSF and the corresponding Met-1 cpds.  
CC The novelty is that hpg-CSF is the prod. of procaryotic or  
eucaryotic expression of an exogenous DNA sequence. The construction  
of hpg-CSF expression vectors is also described in the examples.  
SQ Sequence 17 AA;

Query Match 100.0%; Score 16; DB 1; Length 17;  
Best Local Similarity 25.0%; Pred. No. 1.71e+03;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 2 PLGRASKL 9

QY 2 PXXXXXXL 9

RESULT 4  
ID R30889 standard; peptide; 19 AA.  
AC R30889;  
DT 09-FEB-1993 (first entry)  
DE Cell adhesion polypeptide.  
KW MOLT-4; human; lymphoblastic leukaemia; A375-SM; metastatic;  
melanoma; H1080; fibrosarcoma; LDV; LDL; IDA; inflammatory disease;  
rheumatoid arthritis; asthma; sepsis; graft rejection; reperfusion.  
OS Synthetic.  
PN W09213887-A.  
PD 20-AUG-1992.  
PF 06-FEB-1992; G00226.  
PR 07-FEB-1991; GB-002655.  
PR 08-FEB-1991; GB-002818.  
PA (UYMA-) UNIV VICTORIA MANCHESTER.  
PI Humphries MJ;  
DR WPI: 92-299988/36.  
PT New cell adhesion (poly)peptide(s) modifying cell adhesive  
properties - useful in treating inflammatory conditions e.g.  
rheumatoid arthritis, asthma, inflammatory bowel disease, sepsis,  
etc.  
PS Disclosure; Page 4; 23pp; English.  
CC The peptide is an example of a cell adhesion polypeptide contg. the  
amino sequence X-Asp-Y-(A)n-Phe, where X and Y = Ala, Leu, Ile or  
Val, A = any amino acid and n = 3-10. At least a subsequence of the  
polypeptide is adherent for MOLT-4 human lymphoblastic leukaemia,  
A375-SM human metastatic melanoma or H1080 human fibrosarcoma cells.  
CC The cell adhesion peptides are used to modify or control the  
adhesive properties of cells, e.g. in treatment of inflammatory  
conditions such as rheumatoid arthritis, asthma, sepsis, graft  
rejection, inflammatory bowel disease, reperfusion of cardiac tissue  
after myocardial infarction, and coagulatory disorders. They are  
selective antagonists of cell adhesion, e.g. they promote adhesion  
of the specified cells but inhibit adhesion to the natural adhesion  
protein contg. the adhesive sequence.  
CC See also R26821-30 and R30887-903.  
SQ Sequence 19 AA;

Query Match 100.0%; Score 16; DB 1; Length 19;  
Best Local Similarity 25.0%; Pred. No. 1.71e+03;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 1 PIIDVAPL 8  
QY 2 PXXXXXXL 9

RESULT 5  
ID P70727 standard; Protein; 22 AA.  
AC P70727;  
DT 26-APR-1991 (first entry)  
DE Sequence of portion of human granulocyte colony stimulating  
factor (hpg-CSF).  
KW Haematopoietic disorders; therapy; aplastic anaemia;  
bone marrow transplant; burn wounds; leukaemia.  
OS Homo sapiens.  
FH Key Location/Qualifiers  
FT misc\_difference 12  
FT misc\_difference 14  
PN W08701132-A.  
PD 26-FEB-1987.  
PF 22-AUG-1986; U01708.  
PR 23-AUG-1985; US-768954.  
PR 23-AUG-1985; US-768959.  
PR 03-MAR-1986; US-835548.  
PA (KIRI-) KIRIN-AMGEN INC.  
PA (KIRI) KIRIN-AMGEN INC.  
PI Souza LM;  
DR WPI: 87-064855/09.

\*\*\*\*\*

MPERCH\_PP protein - protein database search, using Smith-Waterman algorithm

(TM)

\*\*\*\*\*

Release 3.1A John F. Collins, BioComputing Research Unit.

Copyright (c) 1993-1998 University of Edinburgh, U.K.

Distribution rights by Oxford Molecular Ltd

MPERCH\_PP protein - protein database search, using Smith-Waterman algorithm

Run on: Sat Apr 15 02:09:46 2000; MasPar time 3.19 Seconds

Tabular output not generated. 66.894 Million cell updates/sec

Title: >US-08-452-843-29

Description: (1-9) from US08452843.pap

Perfect Score: 16

Sequence: 1 PXXXXXXL 9

Scoring table: PAM 150

Gap 15

Searched: 188963 seqs, 23686106 residues

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database: a-geneseq36

1:geneseqp

Statistics: Mean 7.671; Variance 8.970; scale 0.855

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	16	100.0	11	R06768	Tumour necrosis factor	1.71e+03
2	16	100.0	17	R10804	Peptide with motilin-1	1.71e+03
3	16	100.0	17	P70726	Sequence of N-terminal	1.71e+03
4	16	100.0	19	R30889	Cell adhesion polypept	1.71e+03
5	16	100.0	22	P70727	Sequence of portion of	1.71e+03
6	16	100.0	71	P81221	Methionine-insulin-lik	1.71e+03
7	16	100.0	83	P91090	Sequence of viper veno	1.71e+03
8	16	100.0	108	P83143	Sequence of novel pre-	1.71e+03
9	16	100.0	113	P93152	Sequence of human inte	1.71e+03
10	16	100.0	126	R24108	Humanised anti-Tac ant	1.71e+03
11	16	100.0	130	R22481	Neurotrophic factor 4	1.71e+03
12	16	100.0	134	R22821	R99, K113, E122, M127,	1.71e+03
13	16	100.0	134	R23769	Recombinant light chal	1.71e+03
14	16	100.0	134	P91916	Derived sequence of th	1.71e+03
15	16	100.0	170	P70262	Beta-glucuronidase.	1.71e+03
16	16	100.0	171	R29874	HCV NS4-NS5 peptide N2	1.71e+03
17	16	100.0	174	R15619	HBsAg pre-S region sub	1.71e+03
18	16	100.0	174	P82483	Human G-CSF deriv.	1.71e+03
19	16	100.0	178	R10158	pH15C human cytokine s	1.71e+03
20	16	100.0	189	P40126	Sequence encoded by th	1.71e+03
21	16	100.0	205	R33713	Pg1101.	1.71e+03
22	16	100.0	206	R04127	Stem cell leukaemia (S	1.71e+03
23	16	100.0	210	R22482	Neurotrophic factor 4	1.71e+03

24	16	100.0	233	1	P90424	Human tumour necrosis	1.71e+03
25	16	100.0	249	1	R05858	Stem cell leukaemia (S	1.71e+03
26	16	100.0	254	1	P70263	Protein comprising sig	1.71e+03
27	16	100.0	256	1	R22585	ScFvB18 construct #4.	1.71e+03
28	16	100.0	256	1	R22586	ScFvB18 construct #5.	1.71e+03
29	16	100.0	273	1	R20647	Human mast cell growth	1.71e+03
30	16	100.0	330	1	R29926	Elmeria antigen Exam45	1.71e+03
31	16	100.0	455	1	R20787	TNF-alpha binding prot	1.71e+03
32	16	100.0	477	1	P50586	Spinach ribulose-bipho	1.71e+03
33	16	100.0	483	1	R10582	Mutant alpha-amylase S	1.71e+03
34	16	100.0	485	1	R24016	Fusion protein TNFRFC.	1.71e+03
35	16	100.0	498	1	R19316	Pseudorabies virus gx	1.71e+03
36	16	100.0	521	1	P80807	Sequence of gag protei	1.71e+03
37	16	100.0	522	1	R13917	Delta (466-470) tPA va	1.71e+03
38	16	100.0	559	1	R05518	Brassica microspore-sp	1.71e+03
39	16	100.0	560	1	R12341	T-PA variant contg. fl	1.71e+03
40	16	100.0	754	1	R23172	Mutant thermostable DN	1.71e+03
41	16	100.0	768	1	R29850	HCV NS2-NS4 peptide N2	1.71e+03
42	16	100.0	856	1	P80803	Sequence of env protei	1.71e+03
43	16	100.0	911	1	R10333	Deduced sequence of to	1.71e+03
44	16	100.0	1003	1	R29648	AmEPV Spheroidin prote	1.71e+03
45	16	100.0	1089	1	R08267	Platelet derived growt	1.71e+03

## ALIGNMENTS

### RESULT 1

ID R06768 standard; protein; 11 AA.

AC R06768;

DT 23-OCT-1990 (first entry)

DE Tumour necrosis factor derived peptide.

KW Tumour necrosis factor; TNF; neoplastic disease; autoimmune

KW disease; infection; inflammation; transplant rejection.

OS Synthetic.

PN DE3841759-A.

PD 13-JUN-1990.

PF 12-DEC-1988; 841759.

PR 12-DEC-1988; DE-841759.

PA (BADI) BASF AG.

PI Bohm HJ, Daum L, Haupt A, Schmiel B, Walker N, Zechel JC;

DR WPI; 90-186576/25.

PT New tumour necrosis factor derived peptides- for treating or preventing

PT neoplastic and autoimmune diseases, infection, inflammation and

PT transplant rejection.

PS Example 2: Page 8; 16pp; German.

CC To residue M1 is attached H and to residue L11 OH.

CC This peptide is an example of a highly generic sequence of the

CC formula X-A-Y.

CC A= QRETPEGAELKP, HRETPEWAEAKP, HRETPEAEAPMA, PRDTPGEALKP, PGLQEP,

CC PGQGP or PGLQGP;

CC X= G-NH-CHM-CO, G-NH-CHM-CO-W, G-R-NH-CHM-CO or G-R-NH-CHM-CO-W;

CC Y= Z, NH-CHQ-COZ, V-NH-CHQ-COZ, NH-CHQ-CO-U-Z or V-NH-CHQ-CO-U-Z;

CC G= H or an amino protecting group;

CC Z= OH, NH2 or carboxy protecting group; or G and Z together are a

CC covalent bond or the gp. CO(CH2)ANH; a=1-12;

CC R, U, V and W= peptide chains of 1-4 naturally occurring alpha aminoacids;

CC M and Q= H, isopropyl, CHMe, Et, phenyl, CH(OH), 3-indolyl- or

CC 4-imidazolyl-methyl or (CH2)BT; b=1-6;

CC T= OH, MeO, Mes, isopropyl, phenyl (opt. 4-OH, substd.), HS, NH2, COOH,

CC CONH2, or NH2 C (NH) NH; or

CC M and Q together are (CH2)C-S-S-(CH2)d, (CH2)eCO NH-(CH2)f or

CC (CH)2eNH CO(CH2)gNH CO(CH2) h; c and d=1-4; e and f=1-6; g=1-12.

CC The peptide is a low mol. wt. deriv. of TNF.

CC See also DE3841753-55, DE3841759, DE3841761-64, DE3841767-68.

SQ Sequence 11 AA;

Query Match 100.0%; Score 16; DB 1; Length 11;

Best Local Similarity 25.0%; Pred. No. 1.71e+03;

Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 4 PGLQEPWL 11

Qy 2 PXXXXXXL 9

**THIS PAGE BLANK (USPTO)**

```

DR EMBL; AL021899; CAAL7250.1; -
KW Hypothetical protein.
SQ SEQUENCE 213 AA; 23172 MW; 1272E602 CRC32;

Query Match 100.0%; Score 15; DB 2; Length 213;
Best Local Similarity 25.0%; Pred. No. 5.61e+03;
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 128 PLQPRPTI 135
|
Qy 2 PXXXXXXI 9

RESULT 14
ID 027105 PRELIMINARY; PRT; 223 AA.
AC 027105;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-AUG-1998 (TREMBLrel. 07, Last annotation update)
DE PHOSPHATIDYL SERINE DECARBOXYLASE.
GN MTH1026.
OS Methanobacterium thermoautotrophicum.
OC Archaea; Euryarchaeota; Methanobacteriales; Methanobacteriaceae;
OC Methanobacterium.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DELTA H;
RX MEDLINE; 98037514.
RA SMITH D.R., DOUCETTE-STAMM L.A., DELOUGHERY C., LEE H.-M., DUBOIS J.,
RA ALDREGE T., BASHIRZADEH R., BLAKELY D., COOK R., GILBERT K.,
RA HARRISON D., HOANG L., KEAGLE P., LUMM W., POTHIER B., QIU D.,
RA SPADAFORA R., VICARE R., WANG Y., WIERZBOWSKI J., GIBSON R.,
RA JIWANI N., CARUSO A., BUSH D., SAFER H., PATWELL D., PRABHAKAR S.,
RA MCDUGALL S., SHIMER G., GOYAL A., PIETROVSKI S., CHURCH G.M.,
RA DANIELS C.J., MAO J.-I., RICE P., NOLLING J., REEVE J.N.;
RT "Complete genome sequence of Methanobacterium thermoautotrophicum
deltaH: functional analysis and comparative genomics.";
RL J. Bacteriol. 179:7135-7155(1997).
DR EMBL; AE000875; AAB85522.1; -
SQ SEQUENCE 223 AA; 24896 MW; 5819F28B CRC32;

Query Match 100.0%; Score 15; DB 1; Length 223;
Best Local Similarity 25.0%; Pred. No. 5.61e+03;
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 46 PPSDEDLI 53
|
Qy 2 PXXXXXXI 9

RESULT 15
ID 027104 PRELIMINARY; PRT; 226 AA.
AC 027104;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-AUG-1998 (TREMBLrel. 07, Last annotation update)
DE HYPOTHETICAL 25.9 KD PROTEIN.
GN MTH1025.
OS Methanobacterium thermoautotrophicum.
OC Archaea; Euryarchaeota; Methanobacteriales; Methanobacteriaceae;
OC Methanobacterium.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DELTA H;
RX MEDLINE; 98037514.
RA SMITH D.R., DOUCETTE-STAMM L.A., DELOUGHERY C., LEE H.-M., DUBOIS J.,
RA ALDREGE T., BASHIRZADEH R., BLAKELY D., COOK R., GILBERT K.,
RA HARRISON D., HOANG L., KEAGLE P., LUMM W., POTHIER B., QIU D.,
RA SPADAFORA R., VICARE R., WANG Y., WIERZBOWSKI J., GIBSON R.,
RA JIWANI N., CARUSO A., BUSH D., SAFER H., PATWELL D., PRABHAKAR S.,
RA MCDUGALL S., SHIMER G., GOYAL A., PIETROVSKI S., CHURCH G.M.,
RA DANIELS C.J., MAO J.-I., RICE P., NOLLING J., REEVE J.N.;
RT "Complete genome sequence of Methanobacterium thermoautotrophicum
deltaH: functional analysis and comparative genomics.";
RL J. Bacteriol. 179:7135-7155(1997).
DR EMBL; AE000875; AAB85522.1; -
SQ SEQUENCE 226 AA; 25928 MW; B76E2B39 CRC32;

Query Match 100.0%; Score 15; DB 1; Length 226;
Best Local Similarity 25.0%; Pred. No. 5.61e+03;
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 171 PSVIRNII 178
|
Qy 2 PXXXXXXI 9

Search completed: Sat Apr 15 02:06:25 2000
Job time : 91 secs.

```

Query Match 100.0%; Score 15; DB 1; Length 176;  
Best Local Similarity 25.0%; Pred. No. 5.61e+03;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 101 PLRDFDI 108  
QY 2 PXXXXXXI 9

RESULT 10 PRELIMINARY; PRT; 181 AA.  
ID Q9YDZ0  
AC Q9YDZ0  
DT 01-NOV-1999 (TReMBLrel. 12, Created)  
DT 01-NOV-1999 (TReMBLrel. 12, Last sequence update)  
DT 01-NOV-1999 (TReMBLrel. 12, Last annotation update)  
DE 181AA LONG HYPOTHETICAL PROTEIN.  
GN APE0779.  
OS Aeropyrum pernix.  
OC Archaea; Crenarchaeota; Aeropyrum.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-K1;  
RX MEDLINE: 99310339.  
RA KAWARABAYASI Y., HINO Y., HORIKAWA H., YAMAZAKI S., HAIKAWA Y.,  
RA JIN-NO K., TAKAHASHI M., SEKINE M., BABA S., ANKAI A., KOSUGI H.,  
RA HOSOIYAMA A., FUKUI S., NAGAI Y., NISHIJIMA K., NAKAZAWA H.,  
RA TAKAMIYA M., MASUDA S., FUNAHASHI T., TANAKA T., KUDOH Y.,  
RA YAMAZAKI J., KUSHIDA N., OGUCHI A., AOKI K., KUBOTA K., NAKAMURA Y.,  
RA NOMURA N., SAKO Y., KIKUCHI H.;  
RT "Complete genome sequence of an aerobic hyper-thermophilic  
Crenarchaeon, Aeropyrum pernix K1.";  
RL DNA Res. 6:83-101(1999).  
DR EMBL; AP000060; BAA79757.1; -.  
SQ SEQUENCE 181 AA; 20116 MW; 8B8E75A8 CRC32;

Query Match 100.0%; Score 15; DB 1; Length 181;  
Best Local Similarity 25.0%; Pred. No. 5.61e+03;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 3 PILTPFII 10  
QY 2 PXXXXXXI 9

RESULT 11 PRELIMINARY; PRT; 195 AA.  
ID Q9YAK4  
AC Q9YAK4  
DT 01-NOV-1999 (TReMBLrel. 12, Created)  
DT 01-NOV-1999 (TReMBLrel. 12, Last sequence update)  
DT 01-NOV-1999 (TReMBLrel. 12, Last annotation update)  
DE 195AA LONG HYPOTHETICAL PROTEIN.  
GN APE1936.  
OS Aeropyrum pernix.  
OC Archaea; Crenarchaeota; Aeropyrum.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-K1;  
RX MEDLINE: 99310339.  
RA KAWARABAYASI Y., HINO Y., HORIKAWA H., YAMAZAKI S., HAIKAWA Y.,  
RA JIN-NO K., TAKAHASHI M., SEKINE M., BABA S., ANKAI A., KOSUGI H.,  
RA HOSOIYAMA A., FUKUI S., NAGAI Y., NISHIJIMA K., NAKAZAWA H.,  
RA TAKAMIYA M., MASUDA S., FUNAHASHI T., TANAKA T., KUDOH Y.,  
RA YAMAZAKI J., KUSHIDA N., OGUCHI A., AOKI K., KUBOTA K., NAKAMURA Y.,  
RA NOMURA N., SAKO Y., KIKUCHI H.;  
RT "Complete genome sequence of an aerobic hyper-thermophilic  
Crenarchaeon, Aeropyrum pernix K1.";  
RL DNA Res. 6:83-101(1999).  
DR EMBL; AP000062; BAA80945.1; -.  
SQ SEQUENCE 195 AA; 22760 MW; 464BAA61 CRC32;

Query Match 100.0%; Score 15; DB 1; Length 195;  
Best Local Similarity 25.0%; Pred. No. 5.61e+03;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 90 PRVRLRI 97  
QY 2 PXXXXXXI 9

RESULT 12 PRELIMINARY; PRT; 204 AA.  
ID OS8911  
AC OS8911  
DT 01-AUG-1998 (TReMBLrel. 07, Created)  
DT 01-AUG-1998 (TReMBLrel. 07, Last sequence update)  
DT 01-JAN-1999 (TReMBLrel. 09, Last annotation update)  
DE 204AA LONG HYPOTHETICAL PROTEIN.  
GN PH1185.  
OS Pyrococcus horikoshii.  
OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-OT3;  
RX MEDLINE: 98344137.  
RA KAWARABAYASI Y., SAWADA M., HORIKAWA H., HAIKAWA Y., HINO Y.,  
RA YAMAMOTO S., SEKINE M., BABA S., KOSUGI H., HOSOIYAMA A., NAGAI Y.,  
RA SAKAI M., OGURA K., OTUKA R., NAKAZAWA H., TAKAMIYA M., OHFUKU Y.,  
RA FUNAHASHI T., TANAKA T., KUDOH Y., YAMAZAKI J., KUSHIDA N., OGUCHI A.,  
RA AOKI K., NAKAMURA Y., ROBB T.F., HORIKOSHI K., MASUCHI Y., SHIZUYA H.,  
RA KIKUCHI H.;  
RT "Complete sequence and gene organization of the genome of a hyper-  
thermophilic archaeobacterium, Pyrococcus horikoshii OT3.";  
RL DNA Res. 5:55-76(1998).  
DR EMBL; AP000005; BAA30285.1; -.  
SQ SEQUENCE 204 AA; 23451 MW; 63F0629C CRC32;

Query Match 100.0%; Score 15; DB 1; Length 204;  
Best Local Similarity 25.0%; Pred. No. 5.61e+03;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 42 PPEILDKI 49  
QY 2 PXXXXXXI 9

RESULT 13 PRELIMINARY; PRT; 213 AA.  
ID OS3480  
AC OS3480  
DT 01-JUN-1998 (TReMBLrel. 06, Created)  
DT 01-JUN-1998 (TReMBLrel. 06, Last sequence update)  
DT 01-NOV-1998 (TReMBLrel. 08, Last annotation update)  
DE HYPOTHETICAL 23.2 KD PROTEIN.  
GN MT018.23.  
OS Mycobacterium tuberculosis.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-H37RV;  
RA SEEGER K., HARRIS D.;  
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-H37RV;  
RA COLE S.T., PARKHILL J., BARRELL B.G., RAJANDREAM M.A.;  
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-H37RV;  
RA MEDLINE: 96181548.  
RA PHILIPP W.J., POULET S., EIGLMEIER K., PASCOPELLA L.,  
RA BALASUBRAMANIAN V., HEYM B., BERGH S., BLOOM B.R., JACOBS W.R. JR.,  
RA COLE S.T.;  
RT "An integrated map of the genome of the tubercle bacillus,  
Mycobacterium tuberculosis H37RV, and comparison with Mycobacterium  
leprae.";  
RL Proc. Natl. Acad. Sci. U.S.A. 93:3132-3137(1996).

RA YAMAMOTO S., SEKINE M., BABA S., KOSUGI H., HOSOYAMA A., NAGAI Y.,  
 RA SAKAI M., OGURA K., OTUKA R., NAKAZAWA H., TAKAMIYA M., OHFUKU Y.,  
 RA FUNAHASHI T., TANAKA T., KUDOH Y., YAMAZAKI J., KUSHIDA N., OGUCHI A.,  
 RA AOKI K., NAKAMURA Y., ROBB T.F., HORIKOSHI K., MASUCHI Y., SHIZUYA H.,  
 RA KIKUCHI H.;  
 RT "Complete sequence and gene organization of the genome of a hyper-  
 thermophilic archaeobacterium, Pyrococcus horikoshii OT3.";  
 RL DNA Res. 5:55-76(1998).  
 DR EMBL; AP000006; BAA30430.1; -. 514E21A3 CRC32;  
 SQ SEQUENCE 125 AA; 13657 MW; 514E21A3 CRC32;

Query Match 100.0%; Score 15; DB 1; Length 125;  
 Best Local Similarity 25.0%; Pred. No. 5.61e+03;  
 Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 48 PSIGLNNI 55  
 QY 2 PXXXXXXI 9

RESULT 6  
 ID O51431 PRELIMINARY; PRT; 130 AA.  
 AC O51431;  
 DT 01-JUN-1998 (TREMBLrel. 06, Created)  
 DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)  
 DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)  
 DE HYPOTHETICAL 14.9 KD PROTEIN.  
 GN BB0475.  
 OS Borrelia burgdorferi (Lyme disease spirochete).  
 OC Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-ATCC 35210 / B31;  
 RX MEDLINE; 98065943.  
 RA FRASER C.M., CASSENS S., HUANG W.M., SUTTON G.G., CLAYTON R.A.,  
 RA LATHIGRA R., WHITE O., KETCHUM K.A., DODSON R., HICKEY E.K., GWINN M.,  
 RA DOUGHERTY B., TOMB J.-F., FLEISCHMANN R.D., RICHARDSON D.,  
 RA PETERSON J., KERLAVAGE A.R., QUACKENBUSH J., SALZBERG S., HANSON M.,  
 RA VAN VUUT R., PALMER N., ADAMS M.D., GOCAYNE J.D., WEIDMAN J.,  
 RA UTTERBACK T., WATHEY L., MCDONALD L., ARTIACH P., BOWMAN C.,  
 RA GARLAND S., FUJII C., COTTON M.D., HORST K., ROBERTS K., HATCH B.,  
 RA SMITH H.O., VENTER J.C.;  
 RT "Genomic sequence of a Lyme disease spirochaete, Borrelia  
 burgdorferi.";  
 RL Nature 390:580-586(1997).  
 DR EMBL; AF001152; AAC56868.1; -.  
 DR TIGR; BB0475; -.  
 KW Hypothetical protein.  
 SQ SEQUENCE 130 AA; 14886 MW; A248811C CRC32;

Query Match 100.0%; Score 15; DB 2; Length 130;  
 Best Local Similarity 25.0%; Pred. No. 5.61e+03;  
 Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 13 PILVISGI 20  
 QY 2 PXXXXXXI 9

RESULT 7  
 ID O52664 PRELIMINARY; PRT; 131 AA.  
 AC O52664;  
 DT 01-JUN-1998 (TREMBLrel. 06, Created)  
 DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)  
 DT 01-AUG-1998 (TREMBLrel. 07, Last annotation update)  
 DE DSORF-E4.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-EC45;  
 RA WANG Y.-D., ZHAO S., HILL C.W.;

RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF044501; AAC32469.1; -.  
 SQ SEQUENCE 131 AA; 15348 MW; 5D566EFA CRC32;

Query Match 100.0%; Score 15; DB 2; Length 131;  
 Best Local Similarity 25.0%; Pred. No. 5.61e+03;  
 Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 22 PARQEVNI 29  
 QY 2 PXXXXXXI 9

RESULT 8  
 ID P71767 PRELIMINARY; PRT; 144 AA.  
 AC P71767;  
 DT 01-JAN-1998 (TREMBLrel. 05, Created)  
 DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)  
 DT 01-JAN-1998 (TREMBLrel. 05, Last annotation update)  
 DE HYPOTHETICAL 15.2 KD PROTEIN CY277.08 PRECURSOR.  
 GN MTCY277.08.  
 OS Mycobacterium tuberculosis.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-H37RV;  
 RA MURPHY L., HARRIS D., BARRELL B.G., RAJANDREAM M.A., WALSH S.V.;  
 RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.  
 CC -!- SIMILARITY: SOME, TO DROSOPHILA AUBARIA GDH.  
 DR EMBL; 279701; CAB02037.1; -.  
 KW Hypothetical protein; Transmembrane; Signal.  
 FT SIGNAL 1 21 POTENTIAL.  
 FT CHAIN 22 144 HYPOTHETICAL PROTEIN CY277.08.  
 FT TRANSMEM 24 44 POTENTIAL.  
 FT TRANSMEM 45 65 POTENTIAL.  
 SQ SEQUENCE 144 AA; 15159 MW; B6EBCD65 CRC32;

Query Match 100.0%; Score 15; DB 2; Length 144;  
 Best Local Similarity 25.0%; Pred. No. 5.61e+03;  
 Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 2 PVALIWI 9  
 QY 2 PXXXXXXI 9

RESULT 9  
 ID O58119 PRELIMINARY; PRT; 176 AA.  
 AC O58119;  
 DT 01-AUG-1998 (TREMBLrel. 07, Created)  
 DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)  
 DT 01-JAN-1999 (TREMBLrel. 09, Last annotation update)  
 DE 178AA LONG HYPOTHETICAL PROTEIN.  
 GN PH0382.  
 OS Pyrococcus horikoshii.  
 OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-OT3;  
 RX MEDLINE; 98344137.  
 RA KAWARABAYASHI Y., SAWADA M., HORIKAWA H., HAIKAWA Y., HINO Y.,  
 RA YAMAMOTO S., SEKINE M., BABA S., KOSUGI H., HOSOYAMA A., NAGAI Y.,  
 RA SAKAI M., OGURA K., OTUKA R., NAKAZAWA H., TAKAMIYA M., OHFUKU Y.,  
 RA FUNAHASHI T., TANAKA T., KUDOH Y., YAMAZAKI J., KUSHIDA N., OGUCHI A.,  
 RA AOKI K., NAKAMURA Y., ROBB T.F., HORIKOSHI K., MASUCHI Y., SHIZUYA H.,  
 RA KIKUCHI H.;  
 RT "Complete sequence and gene organization of the genome of a hyper-  
 thermophilic archaeobacterium, Pyrococcus horikoshii OT3.";  
 RL DNA Res. 5:55-76(1998).  
 DR EMBL; AP000002; BAA29457.1; -.  
 SQ SEQUENCE 176 AA; 19301 MW; EA92686C CRC32;



```
ID Q58567 PRELIMINARY; PRT; 82 AA.
AC Q58567;
DT 01-JUN-1998 (TREMREL. 06, Created)
DT 01-JUN-1998 (TREMREL. 06, Last sequence update)
DT 01-NOV-1999 (TREMREL. 12, Last annotation update)
DE TUNGSTEN-CONTAINING FORMYLMETHANOFURAN DEHYDROGENASE ISOENZYME II
DE SUBUNIT G (EC 1.2.99.5).
GN FWDG OR MJ1167.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
OC Methanococcus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 96337999.
RA BULT C.J., WHITE O., OLSEN G.J., ZHOU L., FLEISCHMANN R.D.,
RA SUTTON G.G., BLAKE J.A., FITZGERALD L.M., CLAYTON R.A., GOCAYNE J.D.,
RA KERLAVAGE A.R., DOUGHERTY B.A., TOMB J.F., ADAMS M.D., REICH C.I.,
RA OVERBEER R., KIRKNESS E.F., WEINSTOCK K.G., MERRICK J.M., GLODEK A.,
RA SCOTT J.L., GEOHAGAN N.S.M., WEIDMAN J.F., FUHRMANN J.L., NGUYEN D.,
RA UTTERBACK T.R., KELLEY J.M., PETERSON J.D., SADOW P.W., HANNA M.C.,
RA COTTON M.D., ROBERTS K.M., HURST M.A., KAINE B.P., BORODOVSKY M.,
RA KLENK H.-P., FRASER C.M., SMITH H.O., WOESE C.R., VENTER J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii."
RL Science 273:1058-1073(1996).
CC -!- FUNCTION: CATALYZES THE REVERSIBLE OXIDATION OF CO2 AND
CC METHANOFURAN (MPF) TO N-FORMYLMETHANOFURAN (CHO-MPF). THIS ENZYME
CC IS OXYGEN-LABILE. MAY FUNCTION AS AN ELECTRON TRANSFER PROTEIN (BY
CC SIMILARITY).
CC -!- CATALYTIC ACTIVITY: FORMYLMETHANOFURAN + H(2)O + ACCEPTOR = CO(2)
CC + METHANOFURAN + REDUCED ACCEPTOR.
CC -!- COFACTOR: TUNGSTEN. MAY BIND TWO 4FE-4S CLUSTERS (BY SIMILARITY).
CC -!- ENZYME REGULATION: NOT INACTIVATED BY CYANIDE (BY SIMILARITY).
CC -!- PATHWAY: FIRST STEP IN METHANOGENESIS.
CC -!- SUBUNIT: THIS ENZYME IS COMPOSED OF SIX SUBUNITS FWDA, FWDC, FWDD,
CC FWDE, FWDF, AND FWDG.
CC -!- INDUCTION: BY GROWTH ON TUNGSTEN OR MOLYBDENUM UNDER ANAEROBIC
CC CONDITIONS.
CC -!- SIMILARITY: THE IRON-SULFUR CENTERS ARE SIMILAR TO THOSE OF
CC 'BACTERIAL-TYPE' 4FE-4S FERREDOXINS.
DR EMBL; U67558; AAB99169.1; -
DR HGSP; P00198; IFDN.
DR TIGR; MJ1167; -
DR PROSITE; P500198; 4FP4S_FERREDOXIN; 2.
DR PFAM; PF00037; fer4; 1.
DR PRINTS; PR00353; 4FE4SFERDOXIN.
KW Oxidoreductase; Electron transport; Iron-sulfur; 4Fe-4S; Tungsten;
KW Methanogenesis.
FT METAL 13 13 IRON-SULFUR 1 (4FE-4S) (BY SIMILARITY).
FT METAL 16 16 IRON-SULFUR 1 (4FE-4S) (BY SIMILARITY).
FT METAL 19 19 IRON-SULFUR 1 (4FE-4S) (BY SIMILARITY).
FT METAL 23 23 IRON-SULFUR 1 (4FE-4S) (BY SIMILARITY).
FT METAL 60 60 IRON-SULFUR 2 (4FE-4S) (BY SIMILARITY).
FT METAL 63 63 IRON-SULFUR 2 (4FE-4S) (BY SIMILARITY).
FT METAL 66 66 IRON-SULFUR 2 (4FE-4S) (BY SIMILARITY).
FT METAL 70 70 IRON-SULFUR 2 (4FE-4S) (BY SIMILARITY).
SQ SEQUENCE 82 AA; 8797 MW; 9C88B9E6 CRC32;

Query Match 100.0%; Score 15; DB 1; Length 82;
Best Local Similarity 25.0%; Pred. No. 5.61e+03;
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 38 PYSDDDVVI 45
|
QY 2 PXXXXXXI 9

RESULT 3
ID Q49753 PRELIMINARY; PRT; 103 AA.
AC Q49753;
DT 01-NOV-1996 (TREMREL. 01, Created)
DT 01-NOV-1996 (TREMREL. 01, Last sequence update)
DT 01-MAY-1999 (TREMREL. 10, Last annotation update)
```

```
DE HYPOTHETICAL 11.3 KD PROTEIN B1937_F1_22.
GN B1937_F1_22.
OS Mycobacterium leprae.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
RN [1]
RP SEQUENCE FROM N.A.
RA ROBISON K., SMITH D.R.;
RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
DR EMBL; U00016; AAA17157.1; -
DR PFAM; PF00571; CBS; 1.
KW Hypothetical protein.
SQ SEQUENCE 103 AA; 11273 MW; E948AE69 CRC32;

Query Match 100.0%; Score 15; DB 2; Length 103;
Best Local Similarity 25.0%; Pred. No. 5.61e+03;
Matches 2; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Db 19 PSIAIMI 26
|
QY 2 PXXXXXXI 9

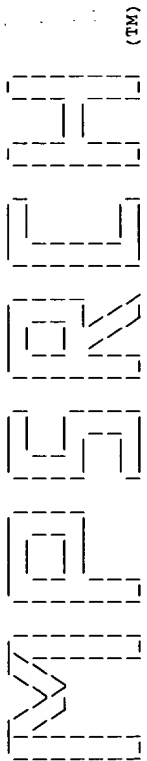
RESULT 4
ID Q9Y9G9 PRELIMINARY; PRT; 109 AA.
AC Q9Y9G9;
DT 01-NOV-1999 (TREMREL. 12, Created)
DT 01-NOV-1999 (TREMREL. 12, Last sequence update)
DT 01-NOV-1999 (TREMREL. 12, Last annotation update)
DE 109AA LONG HYPOTHETICAL PROTEIN.
GN APE2319.
OS Aeropyrum pernix.
OC Archaea; Crenarchaeota; Aeropyrum.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=KI;
RX MEDLINE; 99310339.
RA KAWARABAYASI Y., HINO Y., HORIKAWA H., YAMAZAKI S., HAIKAWA Y.,
RA JIN-NO K., TAKAHASHI M., SEKINE M., BABA S., ANKAI A., KOSUGI H.,
RA HOSOYAMA A., FUKUI S., NAGAI Y., NISHIJIMA K., NAKAZAWA H.,
RA TAKAMIYA M., MASUDA S., FUNAHASHI T., TANAKA T., KUDOH Y.,
RA YAMAZAKI J., KUSHIDA N., OGUCHI A., AOKI K., KUBOTA K., NAKAMURA Y.,
RA NOMURA N., SAKO Y., KIKUCHI H.;
RT "Complete genome sequence of an aerobic hyper-thermophilic
RT crenarchaeon, Aeropyrum pernix KI."
RL DNA Res. 6:83-101(1999).
DR EMBL; AP000064; BAA81331.1; -
SQ SEQUENCE 109 AA; 12423 MW; 73EDBB22 CRC32;

Query Match 100.0%; Score 15; DB 1; Length 109;
Best Local Similarity 25.0%; Pred. No. 5.61e+03;
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 31 PLVQQLPI 38
|
QY 2 PXXXXXXI 9

RESULT 5
ID O59042 PRELIMINARY; PRT; 125 AA.
AC O59042;
DT 01-AUG-1998 (TREMREL. 07, Created)
DT 01-AUG-1998 (TREMREL. 07, Last sequence update)
DT 01-JAN-1999 (TREMREL. 09, Last annotation update)
DE 125AA LONG HYPOTHETICAL PROTEIN.
GN PH1324.
OS Pyrococcus horikoshii.
OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OT3;
RX MEDLINE; 96344137.
RA KAWARABAYASI Y., SAWADA M., HORIKAWA H., HAIKAWA Y., HINO Y.,
```

\*\*\*\*\*



Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Sat Apr 15 02:04:54 2000; Maspar time 14.15 Seconds  
44.091 Million cell updates/sec

Tabular output not generated.

Title: >US-08-452-843-28  
Description: (1-9) from US08452843.peg  
Perfect Score: 15  
Sequence: 1 PXXXXXXI 9

Scoring table: PAM 150  
Gap 15

Searched: 225878 seqs, 69334122 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: spiremb112  
1:sp.archaea 2:sp.bacteria 3:sp.fungi 4:sp.human  
5:sp.invertebrate 6:sp.mammal 7:sp.mhc 8:sp.organelle  
9:sp.phage 10:sp.plant 11:sp.podent 12:sp.unclassified  
13:sp.viridate 14:sp.virus

Statistics: Mean 10.479; Variance 5.866; scale 1.787

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	15	100.0	69	1 Q39827	CONSERVED HYPOTHETICAL	5.61e+03
2	15	100.0	82	1 Q58567	TUNGSTEN-CONTAINING FO	5.61e+03
3	15	100.0	103	2 Q49753	HYPOTHETICAL 11.3 KD P	5.61e+03
4	15	100.0	109	1 Q39959	109AA LONG HYPOTHETICA	5.61e+03
5	15	100.0	125	1 Q59042	125AA LONG HYPOTHETICA	5.61e+03
6	15	100.0	130	2 Q51431	HYPOTHETICAL 14.9 KD P	5.61e+03
7	15	100.0	131	2 Q52664	DSORF-E4	5.61e+03
8	15	100.0	144	2 P71767	HYPOTHETICAL 15.2 KD P	5.61e+03
9	15	100.0	176	1 Q58119	176AA LONG HYPOTHETICA	5.61e+03
10	15	100.0	181	1 Q39D20	181AA LONG HYPOTHETICA	5.61e+03
11	15	100.0	195	1 Q39AK4	195AA LONG HYPOTHETICA	5.61e+03
12	15	100.0	204	1 Q38911	204AA LONG HYPOTHETICA	5.61e+03
13	15	100.0	213	2 Q53480	HYPOTHETICAL 23.2 KD P	5.61e+03
14	15	100.0	223	1 Q27105	PHOSPHATIDYL SERINE DEC	5.61e+03
15	15	100.0	226	1 Q27104	HYPOTHETICAL 25.9 KD P	5.61e+03
16	15	100.0	226	1 Q58371	226AA LONG HYPOTHETICA	5.61e+03
17	15	100.0	243	1 Q31FQ2	243AA LONG HYPOTHETICA	5.61e+03
18	15	100.0	290	1 Q59487	290AA LONG HYPOTHETICA	5.61e+03
19	15	100.0	292	1 Q27819	GLUCOSE-1-PHOSPHATE TH	5.61e+03
20	15	100.0	297	2 Q52689	CCOP.	5.61e+03

21	15	100.0	305	2 O50372	PUTATIVE TRANSPORT PRO	5.61e+03
22	15	100.0	318	2 O53326	HYPOTHETICAL 35.2 KD P	5.61e+03
23	15	100.0	319	1 O26565	HYPOTHETICAL 35.5 KD P	5.61e+03
24	15	100.0	327	1 O58118	327AA LONG HYPOTHETICA	5.61e+03
25	15	100.0	346	1 Q9YDW6	346AA LONG HYPOTHETICA	5.61e+03
26	15	100.0	348	2 O05178	SUGAR TRANSPORTER (GGU	5.61e+03
27	15	100.0	349	1 O59613	349AA LONG HYPOTHETICA	5.61e+03
28	15	100.0	350	2 O52663	CORE PROTEIN (FRAGMENT	5.61e+03
29	15	100.0	355	1 O58566	TUNGSTEN-CONTAINING FO	5.61e+03
30	15	100.0	361	1 Q9YEP3	361AA LONG HYPOTHETICA	5.61e+03
31	15	100.0	374	1 O27079	CARBAMOYL-PHOSPHATE SY	5.61e+03
32	15	100.0	397	1 O58488	397AA LONG HYPOTHETICA	5.61e+03
33	15	100.0	409	2 O54763	HYPOTHETICAL 43.5 KD P	5.61e+03
34	15	100.0	456	1 O29083	SIGNAL-TRANSDUCING HIS	5.61e+03
35	15	100.0	492	1 O28164	GLU-TRNA AMIDOTRANSFER	5.61e+03
36	15	100.0	493	2 O48431	DIHYDROLIPOAMIDE ACETY	5.61e+03
37	15	100.0	531	1 O29488	SIGNAL-TRANSDUCING HIS	5.61e+03
38	15	100.0	537	2 O69807	PUTATIVE TRANSCRIPTION	5.61e+03
39	15	100.0	554	2 O24852	XYLANASE D.	5.61e+03
40	15	100.0	633	2 O50274	ATP SULFURYLASE GTP-BI	5.61e+03
41	15	100.0	663	1 O28469	DNA TOPOISOMERASE I (T	5.61e+03
42	15	100.0	767	2 O51268	CONSERVED HYPOTHETICAL	5.61e+03
43	15	100.0	798	1 O30274	ACETYL-COA DECARBOXYLA	5.61e+03
44	15	100.0	870	1 O27125	DNA-DEPENDENT RNA POLY	5.61e+03
45	15	100.0	1000	2 Q54762	HYPOTHETICAL 114.7 KD	5.61e+03

ALIGNMENTS

RESULT	ID	PRELIMINARY;	PRT;	69 AA.
1	AC O29827			
DT	01-JAN-1998 (TREMBLrel. 05, Created)			
DT	01-JAN-1998 (TREMBLrel. 05, Last sequence update)			
DT	01-AUG-1998 (TREMBLrel. 07, Last annotation update)			
DE	CONSERVED HYPOTHETICAL PROTEIN.			
GN	AF0420.			
OS	Archaeoglobus fulgidus.			
OC	Archaea; Euryarchaeota; Archaeoglobales; Archaeoglobaceae;			
OC	Archaeoglobus.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=VC-16 / DSM 4304 / ATCC 49558;			
RX	MEDLINE; 98049343.			
RA	KLENK H.-P., CLAYTON R.A., TOMB J.-F., WHITE O., NELSON K.E.,			
RA	KETCHUM K.A., DODSON R.J., GWINN M., HICKEY E.K., PETERSON J.D.,			
RA	RICHARDSON D.L., KERLAVAGE A.R., GRAHAM D.E., KYRPIDES N.C.,			
RA	FLEISCHMANN R.D., QUACKENBUSH J., LEE N.H., SUTTON G.G., GILL S.,			
RA	KIRKNESS E.F., DOUGHERTY B.A., MCKENNEY K., ADAMS M.D., LOFTUS B.,			
RA	PETERSON S., REICH C.I., MCNEIL L.K., BADGER J.H., GLODER A., ZHOU L.,			
RA	OVERBECK R., GOCAYNE J.D., WEIDMAN J.F., McDONALD L., UTTERBACK T.,			
RA	COTTON M.D., SPRIGGS T., ARTIACH P., KAINE B.P., SYKES S.M.,			
RA	SADOW P.W., D'ANDREA K.P., BOWMAN C., FUJII C., GARLAND S.A.,			
RA	MASON T.M., OLSEN G.J., FRASER C.M., SMITH H.O., WOESE C.R.,			
RA	VENTER J.C.;			
RT	"The complete genome sequence of the hyperthermophilic, sulphate-			
RT	reducing archaeon Archaeoglobus fulgidus.";			
RL	Nature 390:364-370(1997).			
DR	ENBL; AE001075; AAB90818.1;			
DR	TIGR; AF0420;			
KW	Hypothetical protein.			
SQ	SEQUENCE 69 AA; 8397 MW; AFDDEBC5 CRC32;			

Query Match 100.08; Score 15; DB 1; Length 69;  
Best Local Similarity 25.08; Pred. No. 5.61e+03;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
Db 51 PHHRIIVEI 58  
QY 2 PXXXXXXI 9  
RESULT 2

**THIS PAGE BLANK (USPTO)**

#title Cloning and characterization of the gene for Escherichia coli  
tryptophanyl-transfer ribonucleic acid synthetase.

#cross-references MUID:82075662

#accession I69352

##status translated from GB/EMBL/DDBJ

##molecule\_type DNA

##residues 1-11 ##label RES

##cross-references EMBL:Z28371; NID:G433091; PID:G433092

#accession I54844

##status translated from GB/EMBL/DDBJ

##molecule\_type DNA

##residues 152-171 ##label RE2

##cross-references EMBL:V00370; NID:G43198; PID:G929580

# GENETICS

#gene

#map\_position 74 min

CLASSIFICATION #superfamily tryptophan--trna ligase

KEYWORDS aminocyl-trna synthetase; ATP; ligase; protein biosynthesis

SUMMARY #length 334 #molecular-weight 37438 #checksum 789

Query Match 100.0%; Score 15; DB 1; Length 334;

Best Local Similarity 25.0%; Pred. No. 5.31e+03;

Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 129 PVLMAADI 136

QY 2 PXXXXXXI 9

Search completed: Sat Apr 15 02:03:41 2000

Job time : 16 secs.

```

Db      19 PEIHKPDI 26
      |-----|
Qy      2 PXXXXXXI 9

RESULT  13
ENTRY   QOVZH3      #type complete
TITLE   H3 protein - vaccinia virus (strain WR)
ORGANISM #formal_name vaccinia virus
DATE    31-Mar-1988 #sequence_revision 31-Mar-1988 #text_change
      14-Nov-1997
C24481
ACCESSIONS
REFERENCE A93022
#authors Rosel, J.L.; Earl, P.L.; Weir, J.P.; Moss, B.
#journal J. Virol. (1986) 60:436-449
#title   Conserved TAAATG sequence at the transcriptional and
translational initiation sites of vaccinia virus late genes
deduced by structural and functional analysis of the
HindIII H genome fragment.
#cross-references MUID:87036903
#accession C24481
#molecule_type DNA
#residues 1-324 #label ROS
#cross-references GB:M1309; NID:g335739; PID:g335743
CLASSIFICATION #superfamily vaccinia virus H3 protein
KEYWORDS late protein
SUMMARY #length 324 #molecular-weight 37504 #checksum 7902

Query Match 100.0%; Score 15; DB 1; Length 324;
Best Local Similarity 25.0%; Pred. No. 5.31e-03;
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db      7 PVIVVPVI 14
      |-----|
Qy      2 PXXXXXXI 9

RESULT  14
ENTRY   YWBSF      #type complete
TITLE   tryptophan--trna ligase (EC 6.1.1.2) - Bacillus
steartopherylmophilus
ALTERNATE_NAMES tryptophanyl-trna synthetase
ORGANISM #formal_name Bacillus steartopherylmophilus
DATE    30-Nov-1980 #sequence_revision 30-Sep-1992 #text_change
      30-Jun-1993
ACCESSIONS A26055; A01181
REFERENCE A26055
#authors Barstow, D.A.; Sharman, A.F.; Atkinson, T.; Minton, N.P.
#journal Gene (1986) 46:37-45
#title   Cloning and complete nucleotide sequence of the Bacillus
steartopherylmophilus tryptophanyl trna synthetase gene.
#cross-references MUID:87106841
#accession A26055
#molecule_type DNA
#residues 1-328 #label BAR
#cross-references GB:M14742
REFERENCE A01181
#authors Winter, G.P.; Hartley, B.S.
#journal FEBS Lett. (1977) 80:340-342
#title   The amino acid sequence of tryptophanyl trna synthetase from
Bacillus steartopherylmophilus.
#cross-references MUID:77246821
#accession A01181
#molecule_type protein
#residues 1-31,'2',33-40,'BZ',43-327 #label WIN
GENETICS
#gene trps
CLASSIFICATION #superfamily tryptophan--trna ligase
KEYWORDS aminocacyl-trna synthetase; ATP; ligase; protein biosynthesis
SUMMARY #length 328 #molecular-weight 37117 #checksum 9481

Query Match 100.0%; Score 15; DB 1; Length 328;
Best Local Similarity 25.0%; Pred. No. 5.31e-03;

Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db      51 PHELQONI 58
      |-----|
Qy      2 PXXXXXXI 9

RESULT  15
ENTRY   YWEC      #type complete
TITLE   tryptophan--trna ligase (EC 6.1.1.2) - Escherichia coli
ALTERNATE_NAMES tryptophanyl-trna synthetase
ORGANISM #formal_name Escherichia coli
DATE    13-Jun-1983 #sequence_revision 05-Dec-1997 #text_change
      05-Dec-1997
ACCESSIONS C65133; A01182; S31745; S55290; I69352; I54844
REFERENCE A64720
#authors Blatner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.;
Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.;
Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.;
Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao,
Y.
#journal Science (1997) 277:1453-1462
#title   The complete genome sequence of Escherichia coli K-12.
#cross-references MUID:97428617
#accession C65133
#status nucleic acid sequence not shown; translation not shown
#molecule_type DNA
#residues 1-334 #label BLAT
#cross-references GB:AE000414; GB:U00096; NID:g1789783; PID:g1789786;
#experimental_source strain K-12, substrain MG1655
REFERENCE A92373
#authors Hall, C.V.; van Cleemput, M.; Muench, K.H.; Yanofsky, C.
#journal J. Biol. Chem. (1982) 257:6132-6136
#title   The nucleotide sequence of the structural gene for
Escherichia coli tryptophanyl-trna synthetase.
#cross-references MUID:82189977
#accession A01182
#molecule_type DNA
#residues 1-29,'K',31-325,'O',327-333,'R' #label HAL
#cross-references GB:V00371; NID:g463199; PID:g43200; GB:J01716;
      GB:V00370
REFERENCE A91445
#authors Winter, G.P.; Hartley, B.S.; McLachlan, A.D.; Lee, M.;
Muench, K.H.
#journal FEBS Lett. (1977) 82:348-350
#title   Sequence homology between the tryptophanyl trna synthetase of
Bacillus steartopherylmophilus and Escherichia coli.
#cross-references MUID:78024293
#contents annotation
#note this work confirms parts of the above sequence by amino acid
analysis
REFERENCE S31739
#authors Lyngstadaas, A.; Boye, E.
#submission submitted to the EMBL Data Library, January 1993
#accession S31745
#molecule_type DNA
#residues 1-29,'K',31-325,'O',327-333,'R' #label LYN
#cross-references EMBL:Z19601; NID:g41221; PID:g41228
REFERENCE S55287
#authors Lyngstadaas, A.; Lobner-Olesen, A.; Boye, E.
#journal Mol. Gen. Genet. (1995) 247:546-554
#title   Characterization of three genes in the dam-containing operon
of Escherichia coli.
#cross-references MUID:95327050
#accession S55290
#status preliminary
#molecule_type DNA
#residues 1-29,'K',31-55 #label LV2
#cross-references EMBL:Z19601
      I54844
REFERENCE I54844
#authors Hall, C.V.; Yanofsky, C.
#journal J. Bacteriol. (1981) 148:941-949

```

```
#map_position segment RNA3
CLASSIFICATION #superfamily cucumber mosaic virus coat protein
KEYWORDS       acetylated amino end; coat protein
FEATURE
1
#modified_site acetylated amino end (Met) #status
SUMMARY #length 218 #molecular-weight 24243 #checksum 1039
        experimental
Query Match 100.0%; Score 15; DB 1; Length 218;
Best Local Similarity 25.0%; Pred. No. 5.31e+03;
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 66 PGYFTSI 73
|
QY 2 PXXXXXXI 9

RESULT 10
ENTRY JS0344 #type complete
TITLE tryptophan synthase (EC 4.2.1.20) alpha chain - Lactobacillus
        casei
ORGANISM #formal_name Lactobacillus casei
DATE 31-Mar-1990 #sequence_revision 02-Dec-1994 #text_change
        05-Sep-1997
ACCESSIONS S42347; JS0344
REFERENCE S42342
#authors Natori, Y.; Kano, Y.; Imamoto, F.
#journal J. Biochem. (1990) 107:248-255
#title Nucleotide sequences and genomic constitution of five
        tryptophan genes of Lactobacillus casei.
#accession S42347
#status preliminary
#molecule_type DNA
#residues 1-266 #label NAT
##cross-references EMBL:D00496; NID:9216754; PID:d1000841; PID:g216760
##experimental_source isolate RNL7

GENETICS
#gene trpA
#complex heterodimer of alpha and beta chain
FUNCTION
#description catalyzes the conversion of indolylglycerol phosphate into
        tryptophan and glyceraldehyde phosphate
#pathway tryptophan biosynthesis
CLASSIFICATION #superfamily tryptophan synthase alpha chain; tryptophan
        synthase alpha chain homology
KEYWORDS carbon-oxygen lyase; hydro-lyase; tryptophan biosynthesis
FEATURE
15-240
46
SUMMARY #length 266 #molecular-weight 28724 #checksum 3350
        #domain tryptophan synthase alpha chain homology #label
        TRPA\
        #active_site Glu #status predicted
Query Match 100.0%; Score 15; DB 1; Length 266;
Best Local Similarity 25.0%; Pred. No. 5.31e+03;
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 54 PVADGPVI 61
|
QY 2 PXXXXXXI 9

RESULT 11
ENTRY ERADF3 #type complete
TITLE fiber protein - human adenovirus 3
ORGANISM #formal_name Mastadenovirus h3 #common_name human adenovirus
        3
#note host Homo sapiens (man)
DATE 17-Mar-1987 #sequence_revision 17-Mar-1987 #text_change
        07-Nov-1997
ACCESSIONS A03846
REFERENCE A03846
#authors Signas, C.; Akusjarvi, G.; Pettersson, U.
#journal J. Virol. (1985) 53:672-678

#title Adenovirus 3 fiber polypeptide gene: implications for the
        structure of the fiber protein.
#cross-references MUID:85108162
#accession A03846
#molecule_type DNA
#residues 1-319 #label SIG
##cross-references GB:X0198; GB:M12411; NID:958473; PID:g58474;
        NID:g209926; PID:g209928
#note the authors translated the codon GAA for residue 146 as
        Asp
CLASSIFICATION #superfamily adenovirus fiber protein
KEYWORDS fiber protein; glycoprotein; homotrimer
FEATURE
92,130,188
#binding_site carbohydrate (Asn) (covalent) #status
        predicted
SUMMARY #length 319 #molecular-weight 34815 #checksum 2111
        Query Match 100.0%; Score 15; DB 1; Length 319;
        Best Local Similarity 25.0%; Pred. No. 5.31e+03;
        Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 26 PFINPGFI 33
|
QY 2 PXXXXXXI 9

RESULT 12
ENTRY A41786 #type complete
TITLE mRNA-binding protein p54 - African clawed frog
ORGANISM #formal_name Xenopus laevis #common_name African clawed frog
DATE 04-Mar-1993 #sequence_revision 03-Nov-1995 #text_change
        02-Jul-1998
ACCESSIONS A41786; A36348
REFERENCE A41786
#authors Murray, M.T.; Schiller, D.L.; Franke, W.W.
#journal Proc. Natl. Acad. Sci. U.S.A. (1992) 89:11-15
#title Sequence analysis of cytoplasmic mRNA-binding proteins of
        Xenopus oocytes identifies a family of RNA-binding
        proteins.
#cross-references MUID:92107999
#accession A41786
#molecule_type mRNA
#residues 1-324 #label MUR
##cross-references GB:M80257; NID:g214641; PID:g214642
#note sequence extracted from NCBI backbone (NCBIN:74686,
        NCBIP:74687)
REFERENCE A36348
#authors Murray, M.T.; Krohne, G.; Franke, W.W.
#journal J. Cell Biol. (1991) 112:1-11
#title Different forms of soluble cytoplasmic mRNA binding proteins
        and particles in Xenopus laevis oocytes and embryos.
#cross-references MUID:91093331
#accession A36348
#status nucleic acid sequence not shown; not compared with
        conceptual translation
#molecule_type mRNA
#residues 221-233 #label MU2
#note authors say this sequence was found in the similar p56
        molecule but there are two sequence differences from
        the corresponding region of that protein (see
        accession B38274)
CLASSIFICATION #superfamily Xenopus Y box-binding protein 2; cold shock
        domain homology
KEYWORDS DNA binding; nucleus; oocyte; RNA binding; transcription
        regulation
FEATURE
44-108
SUMMARY #length 324 #molecular-weight 35952 #checksum 3672
        Query Match 100.0%; Score 15; DB 1; Length 324;
        Best Local Similarity 25.0%; Pred. No. 5.31e+03;
        Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
```

Vandenbol, M.; Vannier, F.; Vassarotti, A.; Viari, A.; Wambutt, R.; Wedler, E.; Wedler, H.; Weitzenecker, T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.; Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.

Nature (1997) 390:249-256

The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.

#cross-references MUID:98044033

#accession D69710

##status nucleic acid sequence not shown; translation not shown

##molecule\_type DNA

##residues 1-124 ##label KUN

##cross-references GB:Z99122; GB:Z99123; GB:AL009126; NID:G2636240; PID:el186214; PID:G2636250; NID:G2636029; PID:el184619; PID:G2636238

##experimental\_source strain 168

##comment This protein is involved in the initiation of sporulation.

GENETICS

#gene spoOF

#map\_position 323 (degrees)

CLASSIFICATION #superfamily chemotaxis chey protein; response regulator homology

KEYWORDS phosphoprotein; sporulation

FEATURE 6-115

54 #domain response regulator homology #label RRH\

#binding\_site phosphate (Asp) (covalent) #status predicted

SUMMARY #length 124 #molecular-weight 14228 #checksum 650

Query Match 100.0%; Score 15; DB 1; Length 124;

Best Local Similarity 25.0%; Pred. No. 5.31e+03;

Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 58 PCMDGIEI 65

Qy 2 PXXXXXXI 9

RESULT 7

ENTRY S26725

TITLE Probable transcription termination factor nusA - Thermoplasma acidophilum

ALTERNATE\_NAMES hypothetical protein X (rpoA2 3' region)

ORGANISM #formal\_name Thermoplasma acidophilum

DATE 12-Feb-1993 #sequence\_revision 10-Jul-1998 #text\_change 10-Jul-1998

ACCESSIONS S26725

REFERENCE Klenk, H.P.; Renner, O.; Schwass, V.; Zillig, W.

#authors Nucleic Acids Res. (1992) 20:5226

#journal Nucleotide sequence of the genes encoding the subunits H, B, A' and A' of the DNA-dependent RNA polymerase and the initiator tRNA from Thermoplasma acidophilum.

#accession S26725

##status nucleic acid sequence not shown; translation not shown

##molecule\_type DNA

##residues 1-144 ##label KLE

##cross-references EMBL:X58198; NID:G48089; PID:G48094

##experimental\_source DSM 1728

##note the nucleotide sequence was submitted to the EMBL Data Library, September 1992

CLASSIFICATION #superfamily archaeobacterial probable transcription termination factor nusA

KEYWORDS transcription termination

SUMMARY #length 144 #molecular-weight 16776 #checksum 6417

Query Match 100.0%; Score 15; DB 1; Length 144;

Best Local Similarity 25.0%; Pred. No. 5.31e+03;

Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 110 PEEIGKVI 117

Qy 2 PXXXXXXI 9

RESULT 9

ENTRY VCXVWL

TITLE coat protein - cucumber mosaic virus (strain WL)

ORGANISM #formal\_name cucumber mosaic virus, CMV

#note host Lycopersicon esculentum (tomato)

DATE 05-Apr-1983 #sequence\_revision 30-Jun-1990 #text\_change 22-Jan-1999

ACCESSIONS JA0108; A04216

REFERENCE Davies, C.; Symons, R.H.

#authors Virology (1988) 165:216-224

#journal Further implications for the evolutionary relationships between tripartite plant viruses based on cucumber mosaic virus RNA 3.

#accession JA0108

##status 1-218 ##label DAV

##residues 1-218

##cross-references GB:M21464; NID:G331707; PID:G331709

REFERENCE A61297

#authors Tsunawasa, S.; Narita, K.

#journal J. Biochem. (1982) 92:607-613

#title Micro-identification of amino-terminal acetylaminic acids in proteins.

#contents annotation: acetylation

COMMENT The genome consists of three single-stranded, positive RNAs, designated RNA1, RNA2, and RNA3.

GENETICS

Qy 2 PXXXXXXI 9

RESULT 8

ENTRY VCXVWL

TITLE coat protein - cucumber mosaic virus (strain WL)

ORGANISM #formal\_name cucumber mosaic virus, CMV

#note host Lycopersicon esculentum (tomato)

DATE 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 05-Sep-1997

ACCESSIONS JA0137; JU0087

REFERENCE Ju0087

#authors Quemada, H.; Kearney, C.; Gonsalves, D.; Slightom, J.L.

#journal J. Gen. Virol. (1985) 70:1065-1073

#title Nucleotide sequences of the coat protein genes and flanking regions of cucumber mosaic virus strains C and WL RNA 3.

#cross-references MUID:89279284

#accession JA0137

##molecule\_type genomic RNA

##residues 1-218 ##label QUE

##cross-references EMBL:D00463; NID:G222043; PID:dl000812; PID:G222044

REFERENCE A61297

#authors Tsunawasa, S.; Narita, K.

#journal J. Biochem. (1982) 92:607-613

#title Micro-identification of amino-terminal acetylaminic acids in proteins.

#contents annotation: acetylation

GENETICS

#map\_position segment RNA3

CLASSIFICATION #superfamily cucumber mosaic virus coat protein

KEYWORDS acetylated amino end; coat protein

FEATURE 1

#modified\_site acetylated amino end (Met) #status experimental

SUMMARY #length 218 #molecular-weight 24201 #checksum 1153

Query Match 100.0%; Score 15; DB 1; Length 218;

Best Local Similarity 25.0%; Pred. No. 5.31e+03;

Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 66 PGYFTSI 73

Qy 2 PXXXXXXI 9

RESULT 9

ENTRY VCXVUV

TITLE coat protein - cucumber mosaic virus (strain Q)

ORGANISM #formal\_name cucumber mosaic virus, CMV

#note host Cucumis sativus (cucumber)

DATE 05-Apr-1983 #sequence\_revision 30-Jun-1990 #text\_change 22-Jan-1999

ACCESSIONS JA0108; A04216

REFERENCE Davies, C.; Symons, R.H.

#authors Virology (1988) 165:216-224

#journal Further implications for the evolutionary relationships between tripartite plant viruses based on cucumber mosaic virus RNA 3.

#accession JA0108

##status 1-218 ##label DAV

##residues 1-218

##cross-references GB:M21464; NID:G331707; PID:G331709

REFERENCE A61297

#authors Tsunawasa, S.; Narita, K.

#journal J. Biochem. (1982) 92:607-613

#title Micro-identification of amino-terminal acetylaminic acids in proteins.

#contents annotation: acetylation

COMMENT The genome consists of three single-stranded, positive RNAs, designated RNA1, RNA2, and RNA3.

GENETICS

```

#accession A04426
##molecule_type DNA
##residues 1-88 ##label DUN
REFERENCE
#authors Dunn, J.J.; Studier, F.W.
#journal J. Mol. Biol. (1983) 166:477-535
#title Complete nucleotide sequence of bacteriophage T7 DNA and the
locations of T7 genetic elements.
#cross-references MUID:83241725
#accession S42319
##molecule_type DNA
##residues 1-88 ##label DUW
#cross-references EMBL:V01146; NID:g431187; PID:g15598
##note the authors did not translate the codon for residue 1
GENETICS
#gene 6.7
#map_position 47.23-47.89
CLASSIFICATION
SUMMARY #superfamily phage T7 gene 6.7 protein
#length 88 #molecular-weight 9338 #checksum 3829
Query Match 100.0%; Score 15; DB 1; Length 88;
Best Local Similarity 25.0%; Pred. No. 5.31e-03;
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
Db 10 PKMNTQI 17
QY 2 PXXXXXXI 9
RESULT 5
ENTRY
TITLE TNLJG4
#type complete
trans-activating transcription regulator - simian
immunodeficiency virus (African green monkey isolate)
ORGANISM #formal_name Simian immunodeficiency virus, SIV
DATE 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change
02-Jul-1998
ACCESSIONS E30045
REFERENCE
#authors Fukasawa, M.; Miura, T.; Hasegawa, A.; Morikawa, S.;
Tsujiimoto, H.; Miki, K.; Kitamura, T.; Hayami, M.
#journal Nature (1989) 333:457-461
#title Sequence of simian immunodeficiency virus from African green
monkey, a new member of the HIV/SIV group.
#cross-references MUID:88232906
#accession E30045
##molecule_type DNA
##residues 1-100 ##label FUK
#cross-references EMBL:X07805; NID:g61748; PID:g61753
GENETICS
#gene tat
#introns 73/2
CLASSIFICATION
KEYWORDS #superfamily AIDS trans-activating transcription regulator
#length 100 #molecular-weight 11387 #checksum 9282
SUMMARY
Query Match 100.0%; Score 15; DB 1; Length 100;
Best Local Similarity 25.0%; Pred. No. 5.31e-03;
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
Db 55 PRTRRKKI 62
QY 2 PXXXXXXI 9
RESULT 6
ENTRY
TITLE S2BSOF
#type complete
stage 0 sporulation protein spoOF - Bacillus subtilis
ALTERNATE_NAMES
ORGANISM #formal_name Bacillus subtilis
DATE 28-Dec-1987 #sequence_revision 28-Dec-1987 #text_change
24-Sep-1998
ACCESSIONS A24737; B32354; A23526; S55425; D69710
REFERENCE
A24737
Trach, K.A.; Chapman, J.W.; Piggot, P.J.; Hoch, J.A.
Proc. Natl. Acad. Sci. U.S.A. (1985) 82:7260-7264
Deduced product of the stage 0 sporulation gene spoOF shares
homology with the spoOA, ompR, and sfpA proteins.
#cross-references MUID:86042645
#accession A24737
##molecule_type DNA
##residues 1-124 ##label TRA
#cross-references GB:M11081; NID:g143600; PID:g143601
REFERENCE
#authors Trach, K.; Chapman, J.W.; Piggot, P.; LeCoq, D.; Hoch, J.A.
#journal J. Bacteriol. (1988) 170:4194-4208
#title Complete sequence and transcriptional analysis of the spoOF
region of the Bacillus subtilis chromosome.
#cross-references MUID:88314920
#accession B32354
##molecule_type DNA
##residues 1-123 ##label TR2
REFERENCE
#authors Yoshikawa, H.; Kazami, J.; Yamashita, S.; Chibazakura, T.;
Sone, H.; Kawamura, F.; Oda, M.; Isaka, M.; Kobayashi, Y.;
Saito, H.
#journal Nucleic Acids Res. (1986) 14:1063-1072
#title Revised assignment for the Bacillus subtilis spoOF gene and
its homology with spoOA and with two Escherichia coli
genes.
#cross-references MUID:86120355
#accession A23526
##molecule_type DNA
##residues 1-124 ##label YOS
#cross-references GB:X03497; NID:g40155; PID:g40157
##note both Met-1 and Met-2 are used as initiators when
expressed in E. coli
REFERENCE
#authors Glaser, P.; Danchin, A.
#submission submitted to the EMBL Data Library, May 1995
#description Cloning and sequencing of the Bacillus subtilis chromosomal
region from 320 degrees to 321 degrees.
#accession S55425
##molecule_type DNA
##residues 1-124 ##label GLA
#cross-references EMBL:249782; NID:g853752; PID:g853764
REFERENCE
#authors Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.;
Alloni, G.; Azevedo, V.; Bertero, M.G.; Bessieres, P.;
Bolotin, A.; Borchert, S.; Boriss, R.; Boursier, L.; Brans,
A.; Braun, M.; Brignell, S.C.; Bron, S.; Brouillet, S.;
Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
Choi, S.K.; Codani, J.J.; Connerton, I.F.; Cummings, N.J.;
Daniel, R.A.; Denizot, F.; Devine, K.M.; Duesterhoeft, A.;
Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.;
Fabret, C.; Ferrari, E.; Foulger, D.; Fritz, C.; Fujita,
M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galleron, N.; Ghim,
S.Y.; Glaser, P.; Goffeau, A.; Gollightly, E.J.; Grandi, G.;
Guiseppi, G.; Guy, B.J.; Haga, K.; Haiech, J.; Harwood,
C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.;
Hullo, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karamata, D.;
Kasahara, Y.; Klaerr-Blanchard, M.; Klein, C.; Kobayashi,
Y.; Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.;
Kurita, K.; Lapidus, A.; Lardinois, S.; Lauber, J.;
Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.;
Maueel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno,
M.; Moestl, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly,
V.; Pohl, T.M.; Portetelle, D.; Porwollik, S.; Prescott,
M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro,
A.M.; Prescan, E.; Pujic, P.; Purnelle, B.; Rapoport, G.;
Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.;
Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon, E.;
Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.;
Sekowska, A.; Seror, S.J.; Serror, P.; Shin, B.S.; Soldo,
B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Takahashi, H.;
Takemaru, K.; Takeuchi, M.; Tamakoshi, A.; Tanaka, T.;
Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.;

```



```

ORGANISM #formal_name chloroplast Spinacia oleracea #common_name
spinach
DATE 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change
13-Nov-1998
ACCESSIONS S12198; S18516; JT0211; A27026; S02621; A27243; A30023
REFERENCE #authors Steppuhn, J.; Hermans, J.; Nechushtal, R.; Herrmann, G.S.;
Herrmann, R.G.
#journal Curr. Genet. (1989) 16:99-108
#title Nucleotide sequences of cDNA clones encoding the entire
precursor polypeptide for subunit VI and of the
plasmome-encoded gene for subunit VII of the photosystem I
reaction center from spinach.
#cross-references MUID:90090689
#accession S12198
#status not compared with conceptual translation
#molecule_type mRNA
#residues 1-81 #label STE
#cross-references GB:X16859; NID:g21284; PID:g21285
#accession S18516
#molecule_type protein
#residues 2-7 #label Str2
REFERENCE JT0211
#authors Oh-oka, H.; Takahashi, Y.; Kuriyama, K.; Saeki, K.;
Matsubara, H.
#journal J. Biochem. (1988) 103:962-968
#title The protein responsible for center A/B in spinach photosystem
I: Isolation with iron-sulfur cluster(s) and complete
sequence analysis.
#cross-references MUID:89008208
#accession JT0211
#molecule_type protein
#residues 2-81 #label OHO
REFERENCE A27026
#authors Oh-oka, H.; Takahashi, Y.; Wada, K.; Matsubara, H.; Ohyama,
K.; Ozeki, H.
#journal FEBS Lett. (1987) 218:52-54
#title The 8kDa polypeptide in photosystem I is a probable candidate
of an iron-sulfur center protein coded by the chloroplast
gene frxA.
#accession A27026
#molecule_type protein
#residues 2-30 #label OH2
REFERENCE S02621
#authors Wynn, R.M.; Malkin, R.
#journal FEBS Lett. (1988) 229:293-297
#title Characterization of an isolated chloroplast membrane Fe-S
protein and its identification as the photosystem I Fe-S
(A2)/Fe-S(B) binding protein.
#accession S02621
#status preliminary
#molecule_type protein
#residues 2,'X',4-10,'X',12-13,'X',15-16 #label WYN
GENETICS
#gene psac; frxA
#genome chloroplast
#classification #superfamily ferredoxin 2(4Fe-4S); ferredoxin 2(4Fe-4S)
#keywords 4Fe-4S; chloroplast; electron transfer; iron-sulfur protein;
membrane-associated complex; metalloprotein;
photosynthesis; photosystem I; thylakoid
FEATURE
2-81 #product photosystem I iron-sulfur protein psac #status
experimental #label MAT
4-66 #domain ferredoxin 2(4Fe-4S) homology #label FER
11,14,17,58 #binding_site 4Fe-4S cluster (Cys) (covalent) #status
predicted
21,48,51,54 #binding_site 4Fe-4S cluster (Cys) (covalent) #status
predicted
SUMMARY #length 81 #molecular-weight 9024 #checksum 6599
Query Match 100.0%; Score 15; DB 2; Length 81;
Best Local Similarity 25.0%; Pred. No. 5.31e+03;
Matches 0; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

```

```

Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
Db 22 PTDVLEMI 29
QY 2 PXXXXXXI 9
RESULT 3
ENTRY #type complete
TITLE cytochrome c551 - Pseudomonas fluorescens (biotype C)
ORGANISM #formal_name Pseudomonas fluorescens
DATE 13-Jul-1981 #sequence_revision 13-Jul-1981 #text_change
15-Jan-1999
ACCESSIONS A00092; A90272
REFERENCE A90266
#authors Ambler, R.P.; Wynn, M.
#journal Biochem. J. (1973) 131:485-498
#title The amino acid sequences of cytochromes c-551 from three
species of Pseudomonas.
#cross-references MUID:73224976
#accession A00092
#molecule_type protein
#residues 1-82 #label AMB1
#experimental_source strain C18, ATCC 17400
REFERENCE A90272
#authors Ambler, R.P.
#journal Biochem. J. (1974) 137:13-14
#title The evolutionary stability of cytochrome c-551 in Pseudomonas
aeruginosa and Pseudomonas fluorescens biotype C.
#contents 6 strains, partial sequences
#accession A90272
#molecule_type protein
#residues 1-82 #label AMB2
#note the sequence from strain 50 differs from that shown at
least in having 18-Val, 29-Asp, and 47-Arg; the
sequences from strains 181 and 217 differ in having
1-Asp, 46-Ser, 63-Ala, and 65-Pro; the sequence from
strain 191 differs in having 1-Asp, 46-Asp, and
70-Gln; the sequence from strain 204 differs in having
1-Asp and probably 65-Pro; the sequence from strain
8376 is identical with that shown
GENETICS nlrM
#gene #superfamily cytochrome c6; cytochrome c6 homology
#classification chromoprotein; electron transfer; heme; iron; oxidative
phosphorylation
#keywords #domain cytochrome c6 homology #label CYC\
#binding_site heme (cys) (covalent) #status predicted\
#binding_site heme iron (His, Met) (axial ligands)
#status predicted
SUMMARY #length 82 #molecular-weight 8538 #checksum 8028
Query Match 100.0%; Score 15; DB 1; Length 82;
Best Local Similarity 25.0%; Pred. No. 5.31e+03;
Matches 0; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
Db 11 PCAACHTI 18
QY 2 PXXXXXXI 9
RESULT 4
ENTRY #type complete
TITLE gene 6.7 protein - phase T7
ORGANISM #formal_name phase T7
DATE 13-Jun-1983 #sequence_revision 13-Jun-1983 #text_change
26-Feb-1999
ACCESSIONS A04426; S42319
REFERENCE A94615
#authors Dunn, J.J.; Thompson, K.
#submission submitted to the Nucleic Acid Sequence Database, September
1982

```

\*\*\*\*\*  
W P E R L  
\*\*\*\*\*  
(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (C) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Sat Apr 15 02:03:25 2000; Maspar time 3.15 seconds  
Tabular output not generated.

Title: >US-08-452-843-28  
Description: (1-9) from US08452843.ppe  
Perfect Score: 15  
Sequence: 1 XPXXXXXXI 9

Scoring table: PAM 150  
Gap 15

Searched: 122810 seqs, 40068593 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: pir62  
1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 10.776; Variance 7.464; scale 1.444

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	15	100.0	80	2	A32364	photosystem I iron-su	5.31e+03
2	15	100.0	81	2	S12198	photosystem I iron-su	5.31e+03
3	15	100.0	82	1	CPSP5F	cytochrome c551 - pse	5.31e+03
4	15	100.0	88	1	Q6BP77	gene 6.7 protein - ph	5.31e+03
5	15	100.0	100	1	TNLJG4	trans-activating tran	5.31e+03
6	15	100.0	124	1	S2BSOF	stage 0 sporulation p	5.31e+03
7	15	100.0	144	1	S26725	probable transcrip	5.31e+03
8	15	100.0	218	1	VCVXWL	coat protein - cucumb	5.31e+03
9	15	100.0	218	1	VCVXUV	coat protein - cucumb	5.31e+03
10	15	100.0	266	1	J50344	tryptophan synthase (	5.31e+03
11	15	100.0	319	1	ERADF3	fiber protein - human	5.31e+03
12	15	100.0	324	1	A41786	mRNA-binding protein	5.31e+03
13	15	100.0	324	1	QOVZH3	H3 protein - vaccinia	5.31e+03
14	15	100.0	328	1	YBESF	tryptophan--trna liga	5.31e+03
15	15	100.0	334	1	YWEC	tryptophan--trna liga	5.31e+03
16	15	100.0	344	1	KHPGD	cathepsin D (EC 3.4.2	5.31e+03
17	15	100.0	348	1	PABY	fructose-bisphosphata	5.31e+03
18	15	100.0	358	1	PASPC	fructose-bisphosphata	5.31e+03
19	15	100.0	379	2	I48134	ubiquinol--cytochrome	5.31e+03
20	15	100.0	385	1	S15157	ubiquinol--cytochrome	5.31e+03
21	15	100.0	385	1	CBNC	ubiquinol--cytochrome	5.31e+03
22	15	100.0	387	1	ERADN1	41K fiber protein - h	5.31e+03
23	15	100.0	387	1	CBASN	ubiquinol--cytochrome	5.31e+03

24 15 100.0 389 1 VHIH2E nucleocapsid protein 5.31e+03  
25 15 100.0 402 1 RERTK renin (EC 3.4.23.15) 5.31e+03  
26 15 100.0 407 1 KHRTD cathepsin D (EC 3.4.2 5.31e+03  
27 15 100.0 410 1 KHMSD cathepsin D (EC 3.4.2 5.31e+03  
28 15 100.0 430 2 D70985 probable cytochrome p 5.31e+03  
29 15 100.0 469 1 NMIVN2 exo-alpha-sialidase ( 5.31e+03  
30 15 100.0 475 1 RKSZLN ribulose-bisphosphate 5.31e+03  
31 15 100.0 475 1 RKEGL ribulose-bisphosphate 5.31e+03  
32 15 100.0 515 1 QXZMA NADH dehydrogenase (u 5.31e+03  
33 15 100.0 517 2 S12015 benzoate 4-monooxygen 5.31e+03  
34 15 100.0 532 1 WZBEF5 59K transcription act 5.31e+03  
35 15 100.0 545 2 JX0225 cytochrome P450 CYP10 5.31e+03  
36 15 100.0 547 1 NIZJME nitrogenase molybdenu 5.31e+03  
37 15 100.0 548 1 JS0181 photinus-luciferin 4- 5.31e+03  
38 15 100.0 548 1 S23437 photinus-luciferin 4- 5.31e+03  
39 15 100.0 550 1 A26772 photinus-luciferin 4- 5.31e+03  
40 15 100.0 611 1 S06047 endo-1,4-beta-xylanas 5.31e+03  
41 15 100.0 637 1 QOV211 early transcription f 5.31e+03  
42 15 100.0 724 1 JQ1622 glycoprotein H precu 5.31e+03  
43 15 100.0 925 1 HXAD41 hexon protein - human 5.31e+03  
44 15 100.0 983 1 E45390 env polypotein precu 5.31e+03  
45 15 100.0 990 1 G46335 env polypotein precu 5.31e+03

ALIGNMENTS

RESULT 1  
ENTRY A32364 #type complete  
TITLE Photosystem I iron-sulfur protein - barley chloroplast  
ALTERNATE\_NAMES Photosystem I 9K protein  
ORGANISM #formal\_name Chloroplast Hordeum vulgare #common\_name barley  
DATE 23-Mar-1990 #sequence\_revision 23-Mar-1990 #text\_change 13-Nov-1998

ACCESSIONS A32364  
REFERENCE A32364  
#authors Schellier, H.V.; Svendsen, I.; Moller, B.L.  
#journal Carlsberg Res. Commun. (1989) 54:11-15  
#title Amino acid sequence of the 9-kDa iron-sulfur protein of photosystem I in barley.  
#cross-references MUID:89322553  
#accession A32364  
#status preliminary  
#molecule\_type protein  
#residues 1-80 #label SCH

GENETICS  
#genome chloroplast  
CLASSIFICATION #superfamily ferredoxin 2[4Fe-4S]; ferredoxin 2[4Fe-4S]  
KEYWORDS 4Fe-4S; chloroplast; electron transfer; iron-sulfur protein; membrane-associated complex; metalloprotein; photosynthesis; photosystem I; thylakoid

FEATURE  
3-65 #domain ferredoxin 2[4Fe-4S] homology #label FER\  
10,13,16,57 #binding\_site 4Fe-4S cluster (Cys) (covalent) #status predicted\  
20,47,50,53 #binding\_site 4Fe-4S cluster (Cys) (covalent) #status predicted

SUMMARY  
#length 80 #molecular-weight 8768 #checksum 4299

Query Match 100.0%; Score 15; DB 2; Length 80;  
Best Local Similarity 25.0%; Pred. No. 5.31e+03;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 21 PTDVLEMI 28  
|  
QY 2 PXXXXXXI 9

RESULT 2  
ENTRY S12198 #type complete  
TITLE Photosystem I iron-sulfur protein psaC - spinach chloroplast  
ALTERNATE\_NAMES Photosystem I 9K protein; photosystem I chain VII; photosystem I iron-sulfur protein frxa

```

FT DOMAIN 246 315 SER/THR-RICH.
FT DISULFID 24 53 BY SIMILARITY.
FT DISULFID 57 104 BY SIMILARITY.
FT DISULFID 88 117 BY SIMILARITY.
FT DISULFID 122 163 BY SIMILARITY.
FT DISULFID 149 179 BY SIMILARITY.
FT DISULFID 184 226 BY SIMILARITY.
FT DISULFID 212 242 BY SIMILARITY.
FT CARBOHYD 54 54 POTENTIAL.
FT CARBOHYD 107 107 POTENTIAL.
FT LIPID 312 312 GPI-ANCHOR (BY SIMILARITY).
SQ SEQUENCE 340 AA; 37180 MW; 5CAC18F8 CRC32;

Query Match 100.0%; Score 15; DB 1; Length 340;
Best Local Similarity 25.0%; Pred. No. 4.76e+03;
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 32 PGKDSVI 39
QY 2 PXXXXXXI 9

RESULT 14
ID ATPC-HALVO STANDARD; PRT; 348 AA.
AC Q48330;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE ATP SYNTHASE, SUBUNIT C (EC 3.6.1.34).
GN ATPC.
OS Halobacterium volcanii (Haloferax volcanii).
OC Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae; Haloferax.
RN [1]
RC STRAIN-WR 340;
RX MEDLINE; 95322432.
RA STEINERT K., KROTH-PANCIC P.G., BICKEL-SANDKOTTER S.;
RT "Nucleotide sequence of the ATPase A- and B-subunits of the
RT halophilic archaeobacterium Haloferax volcanii and characterization of
RT the enzyme".
RL Biochim. Biophys. Acta 1249:137-144(1995).
CC -1- SIMILARITY: BELONGS TO A FAMILY THAT GROUPS V-ATPASE SUBUNIT AC39,
CC V-TYPE SODIUM ATPASE NTPC AND ARCHEAL ATPASE SUBUNIT C.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@lsb-sib.ch).
CC -----
DR EMBL; X79516; CAA56049.1; --
KW Hydrolyase; Hydrogen ion transport.
SQ SEQUENCE 348 AA; 38791 MW; E8AA7D86 CRC32;

Query Match 100.0%; Score 15; DB 1; Length 348;
Best Local Similarity 25.0%; Pred. No. 4.76e+03;
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 253 PDELVSKI 260
QY 2 PXXXXXXI 9

RESULT 15
ID 1A43_HUMAN STANDARD; PRT; 365 AA.
AC P30456;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE HLA CLASS I HISTOCOMPATIBILITY ANTIGEN, AW-43 ALPHA CHAIN PRECURSOR.
GN HLA-A OR HLA-A.
```

```

OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A. (A*4301).
RX MEDLINE; 93056508.
RA MADRIGAL J.A., BELICH M.P., HILDEBRAND W.H., BENJAMIN R.J.,
RA LITTLE A.-M., ZEMMOUR J., ENNIS P.D., WARD F.E., PETZL-ERLER M.L.,
RA MARTELL R.W., DU TOIT E.D., PARHAM P.;
RT "Distinctive HLA-A,B antigens of black populations formed by
RT interallelic conversion".
RL J. Immunol. 149:3411-3415(1992).
RN [2]
RP SEQUENCE FROM N.A. (A*4301).
RX MEDLINE; 93235211.
RA MADRIGAL J.A., HILDEBRAND W.H., BELICH M.P., BENJAMIN R.J.,
RA LITTLE A.-M., ZEMMOUR J., ENNIS P.D., WARD F.E., PETZL-ERLER M.L.,
RA DU TOIT E.D., PARHAM P.;
RT "Structural diversity in the HLA-A10 family of alleles: correlations
RT with serology".
RL Tissue Antigens 41:72-80(1993).
CC -1- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO
CC THE IMMUNE SYSTEM.
CC -1- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-
CC MICROGLOBULIN).
CC -1- POLYMORPHISM: THE ONLY ALLELE OF AW-43 KNOWN IS A*4301 WHICH IS
CC SHOWN HERE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@lsb-sib.ch).
CC -----
DR EMBL; X61703; CAA43872.1; --
DR PIR; S16769; S16769.
DR HSSP; P01891; 2HLA.
DR MIM; 142800;
DR PROSITE; PS00290; IG_MHC; 1.
DR PFAM; PF00047; I9; 1.
DR PFAM; PF00129; MHC_I; 1.
KW MHC I; Transmembrane; Glycoprotein; Signal.
FT SIGNAL 1 24
FT CHAIN 25 365 HLA CLASS I HISTOCOMPATIBILITY ANTIGEN,
FT AW-43 ALPHA CHAIN.
FT DOMAIN 25 114 EXTRACELLULAR ALPHA-1.
FT DOMAIN 115 206 EXTRACELLULAR ALPHA-2.
FT DOMAIN 207 298 EXTRACELLULAR ALPHA-3.
FT DOMAIN 299 308 CONNECTING PEPTIDE.
FT TRANSMEM 309 332
FT DOMAIN 333 365 CYTOPLASMIC TAIL.
FT CARBOHYD 110 110 BY SIMILARITY.
FT DISULFID 125 188 BY SIMILARITY.
FT DISULFID 227 283 BY SIMILARITY.
SQ SEQUENCE 365 AA; 41033 MW; 28170E00 CRC32;

Query Match 100.0%; Score 15; DB 1; Length 365;
Best Local Similarity 25.0%; Pred. No. 4.76e+03;
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 304 PTIPVGI 311
QY 2 PXXXXXXI 9

Search completed: Sat Apr 15 02:04:36 2000
Job time : 38 secs.
```

```
Query Match 100.0%; Score 15; DB 1; Length 308;
Best Local Similarity 25.0%; Pred. No. 4.76e+03;
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 158 PSNSLKFI 165
QY 2 PXXXXXXI 9

RESULT 12
ID CYB_POWIS STANDARD; PRT; 308 AA.
AC P16363;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DE CYTOCHROME B (FRAGMENT).
DE MTCTB OR COB OR CYTB.
GN Pomatosomus isidorei (Rufous babbler) (Garritornis isidorei).
OS Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Vertebrata; Archosauria; Aves;
OC Neognathae; Passeriformes; Timaliidae; Pomatostomus.
RN [1]
RN SEQUENCE FROM N.A.
RX MEDLINE; 91288587.
RA EDWARDS S.V., ARCTANDER P., WILSON A.C.;
RT "Mitochondrial resolution of a deep branch in the genealogical tree
RT for perching birds.";
RL Proc. R. Soc. Lond., B, Biol. Sci. 243:99-107(1991).
RN [2]
RN SEQUENCE OF 15-93 FROM N.A.
RX MEDLINE; 89345630.
RA KOCHER T.D., THOMAS W.K., MEYER A., EDWARDS S.V., PAABO S.,
RA VILLALBA F.X., WILSON A.C.;
RT "Dynamics of mitochondrial DNA evolution in animals: amplification
RT and sequencing with conserved primers.";
RL Proc. Natl. Acad. Sci. U.S.A. 86:6196-6200(1989).
CC -1- FUNCTION: COMPONENT OF THE UBIQUINOL-CYTOCHROME C REDUCTASE
CC COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH IS A
CC RESPIRATORY CHAIN THAT GENERATES AN ELECTROCHEMICAL POTENTIAL
CC COUPLED TO ATP SYNTHESIS.
CC -1- CATALYTIC ACTIVITY: QH(2) + 2 FERRICYTOCHROME C - Q +
CC 2 FERROCYTOCHROME C.
CC -1- COFACTOR: TWO HEME GROUPS (B562 AND B566) WHICH ARE NOT COVALENTLY
CC BOUND TO THE PROTEIN.
CC -1- SUBUNIT: THE MAIN SUBUNITS OF COMPLEX B-C1 ARE: CYTOCHROME B,
CC CYTOCHROME C1 AND THE RIESKE PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME B FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X60938; CAA43273.1; -
CC EXBL; M25689; ABA32140.1; -
CC PIR; I33285; I33285.
CC PIR; S22928; S22928.
CC PROSITE; PS00192; CYTOCHROME_B_HEME; 1.
CC PROSITE; PS00193; CYTOCHROME_B_QQ; 1.
CC PFAM; PF00032; cytochrome_b_c1; 1.
CC PFAM; PF00033; cytochrome_b_n; 1.
CC Electron transport; Mitochondrion; Respiratory chain; Transmembrane;
CC Heme.
KW Heme.
FT NON_TER 1 1
FT METAL 51 51 IRON 1 (HEME B562 AXIAL LIGAND).
FT METAL 65 65 IRON 2 (HEME B566 AXIAL LIGAND).
FT METAL 150 150 IRON 2 (HEME B562 AXIAL LIGAND).
FT METAL 164 164 IRON 1 (HEME B566 AXIAL LIGAND).
FT CONFLICT 16 16 G -> A (IN REF. 2).
FT CONFLICT 90 90 A -> T (IN REF. 2).

Query Match 100.0%; Score 15; DB 1; Length 308;
Best Local Similarity 25.0%; Pred. No. 4.76e+03;
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 229 PLATPPHI 236
QY 2 PXXXXXXI 9

RESULT 13
ID DAF_PONPY STANDARD; PRT; 340 AA.
AC P49457;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-FEB-1997 (Rel. 35, Last annotation update)
DE COMPLEMENT DEACY-ACCELERATING FACTOR (CD55) (FRAGMENT).
GN DAF OR CD55.
OS Pongo pygmaeus (Orangutan).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Homnidae; Pongo.
RN [1]
RN SEQUENCE FROM N.A.
RX MEDLINE; 94110622.
RA NICKELLS M.W., ALVAREZ J.I., LUBLIN D.M., ATKINSON J.P.;
RT "Characterization of DAF-2, a high molecular weight form of decay-
RT accelerating factor (DAF; CD55), as a covalently cross-linked dimer
RT of DAF-1.";
RL J. Immunol. 152:676-685(1994).
CC -1- FUNCTION: THIS PROTEIN RECOGNIZES C4B AND C3B FRAGMENTS THAT
CC CONDENSE WITH CELL-SURFACE HYDROXYL OR AMINO GROUPS WHEN NASCENT
CC C4B AND C3B ARE LOCALLY GENERATED DURING C4 AND C3 ACTIVATION.
CC INTERACTION OF DAF WITH CELL-ASSOCIATED C4B AND C3B POLYPEPTIDES
CC INTERFERES WITH THEIR ABILITY TO CATALYZE THE CONVERSION OF C2 AND
CC FACTOR B TO ENZYMICALLY ACTIVE C2A AND BB AND THEREBY PREVENTS
CC THE FORMATION OF C4B2A AND C3BBB, THE AMPLIFICATION CONVEYANCES OF
CC THE COMPLEMENT CASCADE (BY SIMILARITY).
CC -1- SUBUNIT: MONOMER (MAJOR FORM) AND NON-DISULFIDE-LINKED, COVALENT
CC HOMODIMER (MINOR FORM).
CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR.
CC -1- ALTERNATIVE PRODUCTS: TWO FORMS OF DAF (DAF-2, SHOWN HERE, AND
CC DAF-1) ARE PRODUCED BY ALTERNATIVE SPLICING OF THE SAME GENE.
CC -1- DOMAIN: THE FIRST SUSHI DOMAIN (SCR1) IS NOT NECESSARY FOR
CC FUNCTION. SCR2 AND SCR4 PROVIDE THE PROPER CONFORMATION FOR THE
CC ACTIVE SITE ON SCR3 (BY SIMILARITY).
CC -1- PTM: THE SER/THR-RICH DOMAIN IS HEAVILY O-GLYCOSYLATED.
CC -1- SIMILARITY: CONTAINS 4 SUSHI (SCR) REPEATS.
CC -1- SIMILARITY: BELONGS TO THE RECEPTORS OF COMPLEMENT ACTIVATION
CC (RCA) FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; S67775; AAC60609.1; -
CC HSPSP; P08603; IHFI.
CC PFAM; PF00084; sush1; 4.
CC Complement pathway; Plasma; Glycoprotein; Membrane; Repeat;
CC Alternative splicing; GPI-anchor; Sush1.
KW Complement pathway; Plasma; Glycoprotein; Membrane; Repeat;
KW Alternative splicing; GPI-anchor; Sush1.
FT NON_TER 1 1
FT CHAIN <1 312 COMPLEMENT DEACY-ACCELERATING FACTOR.
FT PROPEP 313 340 REMOVED IN MATURE FORM (BY SIMILARITY).
FT DOMAIN <1 243 4 X SUSHI (SCR) REPEATS.
FT REPEAT <1 54 SUSHI 1.
FT REPEAT 56 118 SUSHI 2.
FT REPEAT 121 180 SUSHI 3.
FT REPEAT 183 243 SUSHI 4.
```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; X77731; CAA54787.1; -  
 DR MGD; MGI:102726; DCK.  
 KW Transferase; Kinase; ATP-binding.  
 FT NP\_BIND 28 35 ATP (PROBABLE).  
 SQ SEQUENCE 260 AA; 30367 MW; F2BDDFC CRC32;

Query Match 100.0%; Score 15; DB 1; Length 260;  
 Best Local Similarity 25.0%; Pred. No. 4.76e+03;  
 Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 182 PEKLNRI 189  
 |  
 |  
 Qy 2 PXXXXXXI 9

RESULT 10  
 ID CRTB\_AGRAU STANDARD; PRT; 301 AA.  
 AC P34975;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE PHYTOENE SYNTHASE (EC 2.5.1.-).  
 GN CRTB.  
 OS Agrobacterium aurantiacum.  
 CC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
 CC Rhizobiaceae; Agrobacterium.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 96062243.  
 RA MITSUO N., SATOMI Y., KONDO K., YOKOYAMA A., KAJIWARA S., SAITO T.,  
 RA OHTANI T., MIKI W.;  
 RT "Structure and functional analysis of a marine bacterial carotenoid  
 RT biosynthesis gene cluster and astaxanthin biosynthetic pathway  
 RT proposed at the gene level."  
 RL J. Bacteriol. 177:6575-6584(1995).  
 CC -1- FUNCTION: CATALYSES THE REACTION FROM PREPHYTOENE DIPHOSPHATE  
 CC TO PHYTOENE.  
 CC -1- CATALYTIC ACTIVITY: 2 GERANYLGERANYL DIPHOSPHATE - PYROPHOSPHATE +  
 CC PREPHYTOENE DIPHOSPHATE.  
 CC -1- CATALYTIC ACTIVITY: PREPHYTOENE DIPHOSPHATE - PYROPHOSPHATE +  
 CC PHYTOENE.  
 CC -1- PATHWAY: CAROTENOID BIOSYNTHESIS. INVOLVED IN ASTAXANTHIN  
 CC BIOSYNTHETIC PATHWAY.  
 CC -1- SIMILARITY: BELONGS TO THE PHYTOENE/SQUALENE SYNTHETASE FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; D58420; BAA09595.1; -  
 DR PROSITE; PS01044; SQUALEN\_PHYTOEN\_SYN\_1; 1.  
 DR PROSITE; PS01045; SQUALEN\_PHYTOEN\_SYN\_2; 1.  
 DR PFAM; PF00494; SQS\_PSY\_1;  
 KW Multifunctional enzyme; Carotenoid biosynthesis; Transferase.  
 SQ SEQUENCE 301 AA; 32697 MW; 258DE079 CRC32;

Query Match 100.0%; Score 15; DB 1; Length 301;  
 Best Local Similarity 25.0%; Pred. No. 4.76e+03;  
 Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 229 PPRCAWSI 236  
 |

QY 2 PXXXXXXI 9  
 RESULT 11  
 ID ACPI\_ENTHI STANDARD; PRT; 308 AA.  
 AC P36184; Q24831;  
 DT 01-JUN-1994 (Rel. 29, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE CYSTEINE PROTEINASE ACPI PRECURSOR (EC 3.4.22.-).  
 GN ACPI OR CP3.  
 OS Entamoeba histolytica.  
 CC Eukaryota; Entamoebidae; Entamoeba.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 96422600.  
 RA BRUCHHAUS I., TANNICH E.;  
 RT "A gene highly homologous to ACPI encoding cysteine proteinase 3 in  
 RT Entamoeba histolytica is present and expressed in E. dispar."  
 RL Parasitol. Res. 82:189-192(1996).  
 RN [2]  
 RP SEQUENCE OF 24-308 FROM N.A., AND SEQUENCE OF 93-100.  
 RC STRAIN-HW-1;  
 RX MEDLINE; 93232277.  
 RA REED S., BOUVIER J., POLLACK A.S., ENGEL J.C., BROWN M., HIRATA K.,  
 RA QUE X., EAKIN A., HAGBLOM P., GILLIN F., MCKERROW J.H.;  
 RT "Cloning of a virulence factor of Entamoeba histolytica. Pathogenic  
 RT strains possess a unique cysteine proteinase gene."  
 RL J. Clin. Invest. 91:1532-1540(1993).  
 RN [3]  
 RP SEQUENCE OF 109-274 FROM N.A.  
 RX MEDLINE; 90158686.  
 RA EAKIN A.E., BOUVIER J., SAKANARI J.A., CRAIK C.S., MCKERROW J.H.;  
 RT "Amplification and sequencing of genomic DNA fragments encoding  
 RT cysteine proteases from protozoan parasites."  
 RL Mol. Biochem. Parasitol. 39:1-8(1990).  
 CC -1- FUNCTION: INVOLVED IN PATHOGENICITY. ITS PRESENCE CORRELATES WITH  
 CC INCREASED PROTEINASE EXPRESSION AND ACTIVITY IN PATHOGENIC  
 CC ISOLATES. PROBABLY INVOLVED IN TISSUE INVASION.  
 CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C1; ALSO KNOWN AS THE  
 CC PAPAIN FAMILY OF THIOL PROTEASES.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; X87214; CAA50673.1; -  
 DR EMBL; S58669; AAB26209.1; -  
 DR EMBL; X27307; AAA29094.1; -  
 DR HSSP; P07711; ICJL.  
 DR PROSITE; PS00139; THIOL\_PROTEASE\_CYS; 1.  
 DR PROSITE; PS00639; THIOL\_PROTEASE\_HIS; 1.  
 DR PROSITE; PS00640; THIOL\_PROTEASE\_ASN; 1.  
 DR PFAM; PF00112; Peptidase\_C1; 1  
 KW Hydrolase; Thiol protease; Multigene family; Zymogen; Signal.  
 FT SIGNAL 1 7  
 FT PROPEP 1 92 POTENTIAL.  
 FT CHAIN 93 308 ACTIVATION PEPTIDE.  
 FT ACT\_SITE 115 115 CYSTEINE PROTEINASE ACPI.  
 FT ACT\_SITE 251 251 BY SIMILARITY.  
 FT ACT\_SITE 271 271 BY SIMILARITY.  
 FT ACT\_SITE 112 153 BY SIMILARITY.  
 FT DISULFID 146 186 BY SIMILARITY.  
 FT CONFLICT 61 61 A -> G (IN REF. 2).  
 FT CONFLICT 155 157 GGH -> RG (IN REF. 2).  
 FT CONFLICT 269 270 IR -> VK (IN REF. 3).  
 SQ SEQUENCE 308 AA; 33851 MW; 9415887F CRC32;

```
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
Db 29 POAPELRI 36
QY 2 PXXXXXXI 9

RESULT 7
ID ADHL DROHY STANDARD; PRT; 253 AA.
AC P23236;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE ALCOHOL DEHYDROGENASE 1 (EC 1.1.1.1).
GN ADHL
OS Drosophila hydei (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 91169286.
RA MENOTTI-RAYMOND M., STARNER W.T., SULLIVAN D.T.:
RT "Characterization of the structure and evolution of the Adh region of
RT Drosophila hydei."
RL Genetics 127:355-366(1991).
CC -1- CATALYTIC ACTIVITY: ALCOHOL + NAD(+) - ALDEHYDE OR KETONE + NADH.
CC -1- SUBUNIT: HOMODIMER.
CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC FAMILY (SDR).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announe/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X58694; CAA41540.1; -.
DR PIR; S15712; S15712.
DR FLYBASE; FBgn0012358; DhvY\Adhl.
DR PROSITE; PS00061; ADH_SHORT; 1.
DR PFAM; PF00106; adh_short; 1.
DR PFAM; PF00663; adh_short_C; 1.
KW Oxidoreductase; NAD.
FT INIT_MET 0
FT NP_BIND 9 32 NAD (BY SIMILARITY).
FT ACT_SITE 150 150 BY SIMILARITY.
SQ SEQUENCE 253 AA; 27269 MW; 29C14484 CRC32;

Query Match 100.0%; Score 15; DB 1; Length 253;
Best Local Similarity 25.0%; Pred. No. 4.76e+03;
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 128 PGGVIANI 135
QY 2 PXXXXXXI 9

RESULT 8
ID DCK HUMAN STANDARD; PRT; 260 AA.
AC P27707;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE DEOXYCYTIDINE KINASE (EC 2.7.1.74) (DCK).
GN DCK
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Homiidae; Homo.
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 58-70; 119-127 AND 189-192.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
```

```
RX MEDLINE; 91142207.
RA CHOTTNER E.G., SHEWACH D.S., DATTA N.S., ASHCRAFT E., GRIBBIN D.,
RA GINSBURG D., FOX I.H., MITCHELL B.S.:
RT "Cloning and expression of human deoxycytidine kinase cDNA."
RL Proc. Natl. Acad. Sci. U.S.A. 88:1531-1535(1991).
RN [2]
RP PARTIAL SEQUENCE, AND CHARACTERIZATION.
RX MEDLINE; 91192170.
RA ERIKSSON S., CEDERLUND E., BERGMAN T., JOERNVALL H., BOHMAN C.:
RT "Characterization of human deoxycytidine kinase. Correlation with
RT cDNA sequences."
RL FEBS Lett. 280:363-366(1991).
CC -1- FUNCTION: REQUIRED FOR THE PHOSPHORYLATION OF SEVERAL
CC DEOXYRIBONUCLEOSIDES AND CERTAIN NUCLEOSIDE ANALOGS WIDELY
CC EMPLOYED AS ANTIVIRAL AND CHEMOTHERAPEUTIC AGENTS.
CC -1- CATALYTIC ACTIVITY: NTP + DEOXYCYTIDINE - NDP + CMP.
CC -1- SUBUNIT: HOMODIMER.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- SIMILARITY: BELONGS TO THE DCK/DCK FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announe/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M60527; AAA35752.1; -.
DR PIR; A38585; A38585.
DR PIR; S14321; S14321.
DR MIN; 125450; -.
KW Transferase; Kinase; ATP-binding.
FT NP_BIND 28 35 ATP (PROBABLE).
SQ SEQUENCE 260 AA; 30518 MW; 931CD353 CRC32;

Query Match 100.0%; Score 15; DB 1; Length 260;
Best Local Similarity 25.0%; Pred. No. 4.76e+03;
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 182 PETCLHRI 189
QY 2 PXXXXXXI 9

RESULT 9
ID DCK_MOUSE STANDARD; PRT; 260 AA.
AC P43346;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE DEOXYCYTIDINE KINASE (EC 2.7.1.74) (DCK).
GN DCK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 95014182.
RA KARLSSON A., JOHANSSON M., ERIKSSON S.:
RT "2 cloning and expression of mouse deoxycytidine kinase. Pure
RT recombinant mouse and human enzymes show differences in substrate
RT specificity."
RL J. Biol. Chem. 269:24374-24378(1994).
CC -1- FUNCTION: REQUIRED FOR THE PHOSPHORYLATION OF SEVERAL
CC DEOXYRIBONUCLEOSIDES AND CERTAIN NUCLEOSIDE ANALOGS WIDELY
CC EMPLOYED AS ANTIVIRAL AND CHEMOTHERAPEUTIC AGENTS.
CC -1- CATALYTIC ACTIVITY: NTP + DEOXYCYTIDINE - NDP + CMP.
CC -1- SUBUNIT: HOMODIMER.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- SIMILARITY: BELONGS TO THE DCK/DCK FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
```

products of the carotenoid biosynthesis gene cluster of Rhodobacter capsulatus.";

RT RL Mol. Gen. Genet. 216:254-268(1989).

CC -1- PATHWAY: CAROTENOID AND CHLOROPHYLL BIOSYNTHESIS.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; Z11165; CAA77339.1; -

DR EMBL; X52291; CAA36352.1; -

DR PIR; S04401; S04401.

DR PIR; S17822; S17822.

KW Photosynthesis; Chlorophyll biosynthesis; Carotenoid biosynthesis; Oxidoreductase.

KW OXIDOREDUCTASE.

SQ SEQUENCE 241 AA; 27004 MW; 59085F33 CRC32;

Query Match 100.0%; Score 15; DB 1; Length 241;

Best Local Similarity 25.0%; Pred. No. 4.76e+03;

Matches 2; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Db 18 PWVISOMI 25

OY 2 PXXXXXXI 9

RESULT 6

ID CD8A\_MOUSE STANDARD; PRT; 247 AA.

AC P01731;

DT 21-JUL-1986 (Rel. 01, Created)

DT 21-JUL-1986 (Rel. 01, Last sequence update)

DT 15-JUL-1999 (Rel. 38, Last annotation update)

DE T-CELL SURFACE GLYCOPROTEIN CD8 ALPHA CHAIN PRECURSOR (T-CELL SURFACE GLYCOPROTEIN LYT-2).

GN CD8A OR LYT2 OR LYT-2.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-BALB/C;

RX MEDLINE; 85270477.

RA NAKAUCHI H., NOLAN G.P., HSU C., HUANG H.S., KAVATHAS P., HERZENBERG L.A.;

RT "Molecular cloning of Lyt-2, a membrane glycoprotein marking a subset of mouse T lymphocytes: molecular homology to its human counterpart, Leu-2/78, and to immunoglobulin variable regions.";

RL Proc. Natl. Acad. Sci. U.S.A. 82:5126-5130(1985).

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE; 86079485.

RA ZAMOYSKA R., VOLLMER A.C., SIZER K.C., LIAW C.W., PARNES J.R.;

RT "Two Lyt-2 polypeptides arise from a single gene by alternative splicing patterns of mRNA.";

RL Cell 43:153-163(1985).

RN [3]

RP SEQUENCE FROM N.A.

RX MEDLINE; 87231009.

RA NAKAUCHI H., TAGAWA M., NOLAN G.P., HERZENBERG L.A.;

RT "Isolation and characterization of the gene for the murine T cell differentiation antigen and immunoglobulin-related molecule, Lyt-2.";

RL Nucleic Acids Res. 15:4337-4347(1987).

RN [4]

RP SEQUENCE FROM N.A.

RC STRAIN-C.AKR;

RX MEDLINE; 89006895.

RA YOUNG H.J., HARRISS J.V., GOTTLIEB P.D.;

RT "Nucleotide sequence analysis of the C.AKR Lyt-2a gene: structural polymorphism in alleles encoding the Lyt-2.1 T-cell surface

alloantigen.";

RT RL Immunogenetics 28:345-352(1988).

RN [5]

RP SEQUENCE FROM N.A.

RX MEDLINE; 86252252.

RA LIAW C.W., ZAMOYSKA R., PARNES J.R.;

RT "Structure, sequence, and polymorphism of the Lyt-2 T cell differentiation antigen gene.";

RL J. Immunol. 137:1037-1043(1986).

RN [6]

RP X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS) OF 28-152 IN COMPLEX WITH H-2KB.

RX MEDLINE; 990211475.

RA KERN P.S., TENG M.K., SMOLYAR A., LIU J.H., LIU J., HUSSEY R.E., SPOERL R., CHANG H.-C., REINHERZ E.L., WANG J.-H.;

RT "Structural basis of CD8 coreceptor function revealed by crystallographic analysis of a murine CD8alpha alpha ectodomain fragment in complex with H-2Kb.";

RL Immunity 9:519-530(1998).

CC -1- FUNCTION: IDENTIFIES CYTOTOXIC/SUPPRESSOR T-CELLS THAT INTERACT WITH MHC CLASS I BEARING TARGETS. CD8 IS THOUGHT TO PLAY A ROLE IN THE PROCESS OF T-CELL MEDIATED KILLING. CD8 ALPHA CHAINS BINDS TO CLASS MHC MOLECULES ALPHA-3 DOMAINS.

CC -1- SUBUNIT: IN GENERAL HETERODIMER OF AN ALPHA AND A BETA CHAIN LINKED BY TWO DISULFIDE BONDS. CAN ALSO FORM HOMODIMERS.

CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.

CC -1- ALTERNATIVE PRODUCTS: VARIOUS PATTERNS OF DIFFERENTIAL SPLICING OF CD8 ALPHA TRANSCRIPTS INVOLVE EXCISION OF THE TRANSMEMBRANE OR CYTOPLASMIC DOMAINS.

CC -1- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; M12825; AAA39476.1; -

DR EMBL; M16981; AAA39477.1; ALT\_TERM.

DR EMBL; M12052; AAA39478.1; -

DR EMBL; Y00157; CAA68352.2; -

DR EMBL; M22064; AAA39665.1; -

DR EMBL; M12977; AAA39475.1; -

DR EMBL; M12819; AAA39475.1; JOINED.

DR EMBL; M12975; AAA39475.1; JOINED.

DR EMBL; M12976; AAA39475.1; JOINED.

DR PIR; A01998; RNMST2.

DR PIR; A24784; A24784.

DR PIR; A29523; A29523.

DR PIR; A34954; A34954.

DR PDB; 1BQH; 19-AUG-98.

DR MGD; MGI-88346; CD8A.

DR PFAM; PF00047; Ig; 1.

KW Immunoglobulin domain; Transmembrane; T-cell; Antigen; Glycoprotein;

KW MHC; Signal; Alternative splicing; 3D-structure.

FT SIGNAL 1 27

FT CHAIN 28 247

FT DOMAIN 28 196

FT TRANSMEM 197 217

FT DOMAIN 218 247

FT DISULFID 53 129

FT CARBOHYD 69 97

FT CARBOHYD 97 150

FT CARBOHYD 150 105

FT VARIANT 105 105

FT CONFLICT 81 81

SQ SEQUENCE 247 AA; 27456 MW; 64932EF7 CRC32;

Query Match 100.0%; Score 15; DB 1; Length 247;

Best Local Similarity 25.0%; Pred. No. 4.76e+03;

FT CHAIN 28 247

FT DOMAIN 28 196

FT TRANSMEM 197 217

FT DOMAIN 218 247

FT DISULFID 53 129

FT CARBOHYD 69 97

FT CARBOHYD 97 150

FT CARBOHYD 150 105

FT VARIANT 105 105

FT CONFLICT 81 81

SQ SEQUENCE 247 AA; 27456 MW; 64932EF7 CRC32;

M -> V (IN STRAIN C.AKR).

MISSING (IN REF. 3).



CC -1- SUBUNIT: IN GENERAL HETERODIMER OF AN ALPHA AND A BETA CHAIN  
CC LINKED BY TWO DISULFIDE BONDS. CAN ALSO FORM HOMODIMERS.  
CC -1- SUBCELLULAR LOCATION: TYPE 1 MEMBRANE PROTEIN.  
CC -1- ALTERNATIVE PRODUCTS: VARIOUS PATTERNS OF DIFFERENTIAL SPLICING  
CC OF CD8 ALPHA TRANSCRIPTS INVOLVE EXCISION OF THE TRANSMEMBRANE OR  
CC CYTOPLASMIC DOMAINS.  
CC -1- PTM: ALL OF THE FIVE MOST CARBOXYL-TERMINAL CYSTEINES ARE USED TO  
CC FORM INTER-CHAIN DISULFIDE BONDS IN DIMERS AND HIGHER MULTIMERS,  
CC WHILE THE FOUR AMINO-TERMINAL CYSTEINES ARE NOT (BY SIMILARITY).  
CC -1- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.  
CC -1- DATABASE: NAME=PROW; NOTE=CD guide CD8a entry;  
CC WWW="http://www.ncbi.nlm.nih.gov/prov/cd/cd8alpha.htm".  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; M26315; AAA79217.1; -  
DR EMBL; M26313; AAA79217.1; JOINED.  
DR EMBL; M26314; AAA79217.1; JOINED.  
DR EMBL; M12824; AAA61133.1; -  
DR EMBL; M12828; AAB04637.1; -  
DR EMBL; M27161; AAA59674.1; -  
DR PIR; A01999; RWHU78.  
DR PIR; A2824; A2824.  
DR PIR; JP0105; JP0105.  
DR PIR; A30604; A30604.  
DR PIR; A45888; A45888.  
DR PDB; 1CD8; 31-JAN-94.  
DR MIM; 185910; -  
DR PFAM; PF00047; ig; 1.  
KW Immunoglobulin domain; Transmembrane; Glycoprotein; Phosphorylation;  
KW T-cell; MHC; Signal; 3D-structure; Alternative splicing.  
FT SIGNAL 1 21  
FT CHAIN 22 235 T-CELL SURFACE GLYCOPROTEIN CD8 ALPHA  
FT CHAIN 22 235 CHAIN.  
FT DOMAIN 22 182 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 183 203 POTENTIAL.  
FT DOMAIN 204 235 POTENTIAL.  
FT DOMAIN 22 135 CYTOPLASMIC (POTENTIAL).  
FT DISULFID 43 115 IG-LIKE V-TYPE DOMAIN.  
FT STRAND 24 27  
FT TURN 35 36  
FT TURN 39 45  
FT STRAND 54 59  
FT STRAND 68 73  
FT STRAND 79 80  
FT TURN 82 83  
FT TURN 86 88  
FT STRAND 89 94  
FT TURN 95 96  
FT STRAND 97 102  
FT HELIX 107 109  
FT STRAND 111 119  
FT STRAND 120 121  
FT STRAND 122 125  
FT STRAND 129 131  
FT SEQUENCE 235 AA; CBF991B2 CRC33;

Query Match 100.08; Score 15; DB 1; Length 235;  
Best Local Similarity 25.08; Pred. No. 4.76e+03;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 145 PPTPAPT1 152  
QY 2 PXXXXXXI 9

RESULT 4

ID ATP6\_BACFI STANDARD; PRT; 237 AA.  
AC P22476;  
DT 01-AUG-1991 (Rel. 19, Created)  
DT 01-AUG-1991 (Rel. 19, Last sequence update)  
DT 01-OCT-1994 (Rel. 30, Last annotation update)  
DE ATP SYNTHASE A CHAIN (EC 3.6.1.34) (PROTEIN 6).  
GN ATPB.  
OS Bacillus firmus.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
OC Bacillus/Staphylococcus group; Bacillus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=OF4;  
RX MEDLINE; 92017665.  
RA IVEY D.M., KRULWICH T.A.;  
RT "Organization and nucleotide sequence of the *atp* genes encoding the  
RT *atp* synthase from alkaliphilic *Bacillus firmus* OF4.";  
RL Mol. Gen. Genet. 229:292-300(1991).  
CC -1- FUNCTION: KEY COMPONENT OF THE PROTON CHANNEL; IT MAY PLAY A  
CC DIRECT ROLE IN THE TRANSLOCATION OF PROTONS ACROSS THE MEMBRANE.  
CC -1- SUBUNIT: F-TYPE ATPASES HAVE 2 COMPONENTS, CF(1) - THE CATALYTIC  
CC CORE - AND CF(0) - THE MEMBRANE PROTON CHANNEL. CF(1) HAS FIVE  
CC SUBUNITS: ALPHA(3), BETA(3), GAMMA(1), DELTA(1), EPSILON(1). CF(0)  
CC HAS THREE MAIN SUBUNITS: A, B AND C.  
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. CONTAINS 8  
CC POTENTIAL TRANSMEMBRANE DOMAINS.  
CC -1- SIMILARITY: BELONGS TO THE ATPASE A CHAIN FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; M50117; AAC08038.1; -  
DR PIR; S17720; S17720.  
DR PROSITE; PS00449; ATPASE\_A; 1.  
DR PFAM; PF00119; ATP-synt\_A; 1.  
KW Hydrogen ion transport; CF(0); Transmembrane.  
SQ SEQUENCE 237 AA; 26839 MW; C8C8480A CRC32;  
Query Match 100.0%; Score 15; DB 1; Length 237;  
Best Local Similarity 25.0%; Pred. No. 4.76e+03;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
Db 151 PFLPFI 158  
QY 2 PXXXXXXI 9  
RESULT 5  
ID CRTA\_RHOCA STANDARD; PRT; 241 AA.  
AC P17055;  
DT 01-AUG-1990 (Rel. 15, Created)  
DT 01-MAY-1992 (Rel. 22, Last sequence update)  
DT 01-MAY-1992 (Rel. 22, Last annotation update)  
DE SPHEROIDE MONOOXYGENASE (EC 1.-.-.-).  
GN CRTA.  
OS Rhodobacter capsulatus (Rhodospseudomonas capsulata).  
OC Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;  
OC Rhodobacter.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA BURKE D.H., ALBERTI M., ARMSTRONG G.A., HEARST J.E.;  
RL Submitted (NOV-1991) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP PRELIMINARY SEQUENCE FROM N.A.  
RC STRAIN=SB1003 AND BEC404;  
RX MEDLINE; 89313663.  
RA ARMSTRONG G.A., ALBERTI M., LEACH F., HEARST J.E.;  
RT "Nucleotide sequence, organization, and nature of the protein



DR HSP: P02901; IACP.  
DR MENDEL; 468; HORVU; AC11.1.  
DR PROSITE; PS00012; PHOSPHOPANTHETHEINE; 1.  
DR PROSITE; PS00075; ACP DOMAIN; 1.  
DR PFAM; PF00550; PP-binding; 1.  
KW Fatty acid biosynthesis; Phosphopantetheine; Chloroplast;  
KW Transit peptide; Multigene family.  
FT TRANSIT 1 59  
FT CHAIN 60 149 ACYL CARRIER PROTEIN I.  
FT BINDING 104 104 PHOSPHOPANTHETHEINE.  
SQ SEQUENCE 149 AA; 15974 MW; 198DD03 CRC32;

Query Match 100.0%; Score 15; DB 1; Length 149;  
Best Local Similarity 25.0%; Pred. No. 4.76e+03;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 50 PSSLRFKI 57  
QY 2 PXXXXXXI 9

RESULT 2 STANDARD; PRT; 214 AA.

AC P87421;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-DEC-1999 (Rel. 39, Last sequence update)  
DT 15-DEC-1999 (Rel. 39, Last annotation update)  
DE CYTOCHROME B (FRAGMENT).  
GN MTCYB OR COB OR CYTB.  
OS Atractaspis micropholis.  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Atractaspidae;  
OC Squamata; Scleroglossa; Serpentes; Colubroidea; Atractaspidae;  
OC Atractaspis.  
[1]  
RN SEQUENCE FROM N.A.  
RX MEDLINE; 98334559.  
RA VIDAL N., LECOINTRE G.;  
RT "Weighting and congruence: a case study based on three mitochondrial  
genes in pitvipers";  
RL Mol. Phylogenet. Evol. 9:366-374(1998).  
[2]  
RN SEQUENCE OF 1-132 FROM N.A.  
RA VIDAL N., LECOINTRE G., VIE J.-C., GASC J.-P.;  
RT "Molecular systematics of pitvipers: paraphyly of the Bothrops  
complex";  
RL C. R. Acad. Sci., III, Sci. Vie 320:95-101(1997).  
CC -1- FUNCTION: COMPONENT OF THE UBIQUINOL-CYTOCHROME C REDUCTASE  
COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH IS A  
RESPIRATORY CHAIN THAT GENERATES AN ELECTROCHEMICAL POTENTIAL  
COUPLED TO ATP SYNTHESIS.  
CC -1- CATALYTIC ACTIVITY: QH(2) + 2 FERRICYTOCHROME C = Q +  
2 FERROCYTOCHROME C.  
CC -1- COFACTOR: TWO HEME GROUPS (B562 AND B566) WHICH ARE NOT COVALENTLY  
BOUND TO THE PROTEIN.  
CC -1- SUBUNIT: THE MAIN SUBUNITS OF COMPLEX B-C1 ARE: CYTOCHROME B,  
CYTOCHROME C1 AND THE RIESKE PROTEIN.  
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME B FAMILY.

CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AF039261; AAC33538.1;  
DR PROSITE; PS00192; CYTOCHROME B\_HEME; FALSE\_NEG.  
DR PROSITE; PS00193; CYTOCHROME\_B\_OO; PARTIAL.  
DR PFAM; PF00033; cytochrome\_b\_n; 1.  
KW Electron transport; Mitochondrion; Respiratory chain; Transmembrane;  
Heme.

FT NON\_TER 1 1  
FT METAL 81  
FT METAL 95  
FT NON\_TER 214  
SQ SEQUENCE 214 AA; 24059 MW; 48C843A6 CRC32;

Query Match 100.0%; Score 15; DB 1; Length 214;  
Best Local Similarity 25.0%; Pred. No. 4.76e+03;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 206 PLGTNSDI 213  
QY 2 PXXXXXXI 9

RESULT 3 STANDARD; PRT; 235 AA.

AC P01732;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 21-JUL-1986 (Rel. 01, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE T-CELL SURFACE GLYCOPROTEIN CD8 ALPHA CHAIN PRECURSOR (T-LYMPHOCYTE  
DE DIFFERENTIATION ANTIGEN T8/LEU-2).  
GN CD8A OR MAL.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.  
[1]  
RN SEQUENCE FROM N.A.  
RX MEDLINE; 85099337.  
RA LITTMAN D.R., THOMAS Y., MADDON P.J., CHESSE L., AXEL R.;  
RT "The isolation and sequence of the gene encoding T8: a molecule  
defining functional classes of T lymphocytes";  
RL Cell 40:237-246(1985).  
[2]  
RN SEQUENCE FROM N.A.  
RX MEDLINE; 86103103.  
RA PARNES J.R., SIZER K.C., SUKHATME V.P., HUNKAPILLER T.;  
RT "Structure of Leu-2/T8 as deduced from the sequence of a cDNA clone";  
RL Behring Inst. Mitt. 77:48-55(1985).  
[3]  
RN SEQUENCE FROM N.A.  
RX MEDLINE; 85124610.  
RA SUKHATME V.P., SIZER K.C., VOLLMER A.C., HUNKAPILLER T.,  
RA PARNES J.R.;  
RT "The T cell differentiation antigen Leu-2/T8 is homologous to  
immunoglobulin and T cell receptor variable regions";  
RL Cell 40:591-597(1985).  
[4]  
RN SEQUENCE FROM N.A.  
RX MEDLINE; 90035142.  
RA NAKAYAMA K.-I., TOKITO S., OKUMURA K., NAKAUCHI H.;  
RT "Structure and expression of the gene encoding CD8 alpha chain (Leu-  
2/T8)";  
RL Immunogenetics 30:393-397(1989).  
[5]  
RN SEQUENCE FROM N.A.  
RX MEDLINE; 89215302.  
RA NORBERT A.M., LONBERG N., LACY E., LITTMAN D.R.;  
RT "Alternatively spliced mRNA encodes a secreted form of human CD8  
alpha. Characterization of the human CD8 alpha gene";  
RL J. Immunol. 142:3312-3319(1989).  
[6]  
RN X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS) OF 22-135.  
RX MEDLINE; 92191292.  
RA LEAHY D.J., AXEL R., HENDRICKSON W.A.;  
RT "Crystal structure of a soluble form of the human T cell coreceptor  
CD8 at 2.6-A resolution";  
RL Cell 68:1145-1162(1992).  
CC -1- FUNCTION: IDENTIFIES CYTOTOXIC/SUPPRESSOR T-CELLS THAT INTERACT  
WITH MHC CLASS I BEARING TARGETS. CD8 IS THOUGHT TO PLAY A ROLE IN  
THE PROCESS OF T-CELL MEDIATED KILLING. CD8 ALPHA CHAINS BINDS TO  
CLASS MHC MOLECULES ALPHA-3 DOMAINS.

\*\*\*\*\*

(TM)

\*\*\*\*\*

Release 3.1A John F. Collins, BioComputing Research Unit.

Copyright (c) 1993-1998 University of Edinburgh, U.K.

Distribution rights by Oxford Molecular Ltd

MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Sat Apr 15 02:03:58 2000; Maspar time 3.14 Seconds

Tabular output not generated. 85.467 Million cell updates/sec

Title: &gt;US-08-452-843-28

Description: (1-9) from US08452843.pgp

Perfect Score: 15

Sequence: 1 XPXXXXXXI 9

Scoring table: PAM 150

Gap 15

Searched: 82229 seqs, 29864866 residues

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database: swiss-prot38

1:swissprot

Statistics: Mean 11.429; Variance 7.055; scale 1.620

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	15	100.0	149	1	ACPL_HORVU ACYL CARRIER PROTEIN I	4.76e+03
2	15	100.0	214	1	CYB_ATRMI CYTOCHROME B (FRAGMENT	4.76e+03
3	15	100.0	235	1	CDBA_HUMAN T-CELL SURFACE GLYCOP	4.76e+03
4	15	100.0	237	1	ATP6_BACFI ATP SYNTHASE A CHAIN (	4.76e+03
5	15	100.0	241	1	CRTA_RHOCA SPHEROIDENE MONOOXYGEN	4.76e+03
6	15	100.0	247	1	CDBA_MOUSE T-CELL SURFACE GLYCOP	4.76e+03
7	15	100.0	253	1	ADHL_DROHY ALCOHOL DEHYDROGENASE	4.76e+03
8	15	100.0	260	1	DEK_HUMAN DEOXYCYTIDINE KINASE (	4.76e+03
9	15	100.0	260	1	DEK_MOUSE DEOXYCYTIDINE KINASE (	4.76e+03
10	15	100.0	301	1	CRTB_AGRAU PHYTOENE SYNTHASE (EC	4.76e+03
11	15	100.0	308	1	ACPL_ENTHI CYSTEINE PROTEINASE AC	4.76e+03
12	15	100.0	308	1	CYB_POMIS CYTOCHROME B (FRAGMENT	4.76e+03
13	15	100.0	340	1	DAF_PONPY COMPLEMENT DECAV-ACCEL	4.76e+03
14	15	100.0	348	1	ATPC_HALVO ATP SYNTHASE, SUBUNIT	4.76e+03
15	15	100.0	365	1	IA43_HUMAN HLA CLASS I HISTOCOMPA	4.76e+03
16	15	100.0	375	1	ACT_SCHPO ACTIN.	4.76e+03
17	15	100.0	377	1	ACT_COLSC ACTIN.	4.76e+03
18	15	100.0	378	1	ACT_SCHDU ACTIN.	4.76e+03
19	15	100.0	381	1	DAF_HUMAN COMPLEMENT DECAV-ACCEL	4.76e+03
20	15	100.0	390	1	SHIB_HUMAN 5-HYDROXYTRYPTAMINE 1B	4.76e+03
21	15	100.0	399	1	ATPC_METUA ATP SYNTHASE, SUBUNIT	4.76e+03
22	15	100.0	456	1	ACHA_BRARE ACETYLCHOLINE RECEPTOR	4.76e+03
23	15	100.0	456	1	ACHA_CHICK ACETYLCHOLINE RECEPTOR	4.76e+03

24 15 100.0 461 1 D15K\_MOUSE D15K21 PROTEIN (FRAGME 4.76e+03

25 15 100.0 462 1 ANX7\_DICDI ANNEXIN VII (SYNEXIN). 4.76e+03

26 15 100.0 466 1 DCEA\_ECOLI GLUTAMATE DECARBOXYLAS 4.76e+03

27 15 100.0 466 1 DCEB\_ECOLI GLUTAMATE DECARBOXYLAS 4.76e+03

28 15 100.0 479 1 AACA\_STRAU BIFUNCTIONAL AAC/APH ( 4.76e+03

29 15 100.0 503 1 ATPA\_SYNY3 ATP SYNTHASE ALPHA CHA 4.76e+03

30 15 100.0 511 1 CP45\_RABIT CYTOCHROME P450 4A5 PR 4.76e+03

31 15 100.0 514 1 ATPA\_THIFE ATP SYNTHASE ALPHA CHA 4.76e+03

32 15 100.0 578 1 ATPA\_METMA ATP SYNTHASE ALPHA CHA 4.76e+03

33 15 100.0 591 1 COXN\_BRAJA ALTERNATIVE CYTOCHROME 4.76e+03

34 15 100.0 631 1 DMK\_MOUSE MYOTONIN-PROTEIN KINAS 4.76e+03

35 15 100.0 702 1 ATIL\_VARY 81 KD A-TYPE INCLUSION 4.76e+03

36 15 100.0 717 1 CYG5\_HUMAN GUANYLATE CYCLASE SOLU 4.76e+03

37 15 100.0 724 1 ATIL\_VACCV 94 KD A-TYPE INCLUSION 4.76e+03

38 15 100.0 732 1 CYG4\_HUMAN GUANYLATE CYCLASE SOLU 4.76e+03

39 15 100.0 761 1 CTPA\_MYCTU CATION-TRANSPORTING P- 4.76e+03

40 15 100.0 986 1 CIGR\_ARBPV RESACT RECEPTOR PRECU 4.76e+03

41 15 100.0 1001 1 ATCA\_RABIT CALCIUM-TRANSPORTING A 4.76e+03

42 15 100.0 1048 1 ANGR\_VIBAN ANGR PROTEIN. 4.76e+03

43 15 100.0 1125 1 CYGS\_STRPU SPERACT RECEPTOR PRECU 4.76e+03

44 15 100.0 1472 1 ATC9\_YEAST PROBABLE CALCIUM-TRANS 4.76e+03

45 15 100.0 2032 1 CTOG\_HUMAN CH-TOG PROTEIN (COLONI 4.76e+03

## ALIGNMENTS

RESULT 1

AC ACPL\_HORVU STANDARD; PRT; 149 AA.

DT 21-JUL-1986 (Rel. 01, Created)

DT 01-APR-1990 (Rel. 14, Last sequence update)

DT 15-DEC-1998 (Rel. 37, Last annotation update)

DE ACYL CARRIER PROTEIN I PRECURSOR (ACP I).

GN ACL1.1.

OS Hordeum vulgare (Barley).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC eukaryotes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;

OC Poaceae; Hordeum.

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE; 92049248.

RA HANSEN L., VON WEITSTEIN-KNOWLES P.;

RT "The barley genes Acl1 and Acl3 encoding acyl carrier proteins I and

RT III are located on different chromosomes.";

RL Mol. Gen. Genet. 229:467-478(1991).

RN [2]

RP SEQUENCE FROM N.A.

RA HANSEN L.;

RT "Three cDNA clones for barley leaf acyl carrier proteins I and III.";

RL Carlsberg Res. Commun. 52:381-392(1987).

RN [3]

RP SEQUENCE OF 60-131.

RA HOJ P.B., SVENDSEN I.;

RT "Barley acyl carrier protein: its amino acid sequence and assay using

RT purified malonyl-CoA:ACP transacylase.";

RL Carlsberg Res. Commun. 48:285-305(1983).

CC -!- FUNCTION: THIS PROTEIN IS THE CARRIER OF THE GROWING FATTY ACID

CC CHAIN IN FATTY ACID BIOSYNTHESIS.

CC -!- SUBCELLULAR LOCATION: CHLOROPLAST.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC

CC ENBL; M24425; AAA32923.1; -

CC ENBL; M58753; AAA32920.1; -

CC PIR; A03399; AYBH.

CC PIR; A29638; A29638.

CC PIR; S17927; S17927.

KW cDNA; clone pz130; anthesis; tomato; ovary; integumen; outer pericarp;  
 KW fruit; development; transcription; initiation; region; modulation;  
 KW ovary-specific; endogenous; fruit product; exogenous; phenotype.  
 OS Lycopersicon esculentum.

FH Key Location/Qualifiers

FT misc\_difference 119 /note= "Nonsense codon"  
 FT misc\_difference 120 /note= "Nonsense codon"  
 FT misc\_difference 126 /note= "Nonsense codon"  
 FT misc\_difference 129 /note= "Nonsense codon"  
 FT misc\_difference 146 /note= "Nonsense codon"  
 FT misc\_difference 152 /note= "Nonsense codon"  
 FT misc\_difference 160 /note= "Nonsense codon"  
 FT misc\_difference 164 /note= "Nonsense codon"  
 FT misc\_difference 181 /note= "Nonsense codon"  
 FT /note= "Nonsense codon"

US5175095-A.

29-DEC-1992.

PD 19-JUL-1989; 382518.

PR 19-JUL-1989; US-382518.

PR 17-JUL-1990; US-554195.

PA (CALT) CALGENE INC.

PI Houck CM, Martineau BM;

DR WPI; 93-026940/03.

DR N-PSDB; Q34940.

PT DNA constructs contg. tomato pz130 transcriptional initiation

PT region - useful for modulation of endogenous fruit prods. and for

PT prodn. of exogenous prods.

PS Disclosure: Fig 1; 18pp; English.

CC The sequence represents the polypeptide of cDNA clone pz130. This

CC sequence is expressed during the early stages of anthesis in tomatoes.

CC The message is expressed in ovary integumen and ovary outer pericarp

CC tissue. It is not readily detectable in other tissues or at other

CC stages of fruit development. The transcription initiation region

CC associated with this gene is therefore considered to be ovary-

CC specific. The actual function of the pz130 polypeptide is unknown.

CC The transcription initiation region can be used for modulation of

CC endogenous fruit products, for production of exogenous products and

CC for modification of the phenotype of fruit and fruit products.

SC Sequence 188 AA;

Query Match 100.0%; Score 15; DB 1; Length 188;

Best Local Similarity 37.5%; Pred. No. 4.03e+03;

Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Db 155 PIVAXHI 162

Oy 2 PXXXXXXI 9

RESULT 15

ID R07596 standard; protein; 191 AA.

AC R07596.

DT 20-DEC-1990 (first entry)

DE Animal somatotropin analogue #3.

KW bovine somatotropin; milk production.

OS Bos taurus.

PN W09008164-A.

PD 26-JUL-1990.

PF 11-DEC-1989; U05447.

PR 19-JAN-1989; US-299107.

PA (UPJO) UPJOHN CO.

PI Garlick RL, Lyle SB, Mott JE;

DR WPI; 90-254016/33

PT Animal somatotropin analogues - having substitute for asparagine

PT at position 99 for improved storage stability and enhanced

PT bio-activity.  
 PS Claim 1; Page 20; 25pp; English.  
 CC Sequence corresponds to the sequence of bovine somatotropin with a  
 CC substitution at posn. 99; Gly replaces Asn. The substitution is  
 CC introduced by a site-directed mutagenic technique using an oligomer  
 CC denoted CST-84. The mutated cDNA  
 CC sequence encoding this analogue was excised from the parental  
 CC vector and cloned into the pURA-m4 vector and then transformed into  
 CC fermentation expression strain Bst-1C. The transformants were  
 CC cultured to produce the somatotropin analogue which can be used to  
 CC increase milk production in cows. It also has improved liquid  
 CC storage stability and uniform potency.  
 CC See also R06384, R07595, R07597-R07598 and R07635-R07636.  
 SQ Sequence 191 AA;

Query Match 100.0%; Score 15; DB 1; Length 191;

Best Local Similarity 25.0%; Pred. No. 4.03e+03;

Matches 2; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Db 38 PEGORYSI 45

Oy 2 PXXXXXXI 9

Search completed: Sat Apr 15 02:03:07 2000

Job time : 37 secs.

CC for the human CD3 antigen and due to the lack of immunological  
 CC response caused by the synthetic CDR's the ligand can be considered to  
 CC be humanised. This ligand can be used to manufacture medicaments  
 CC for use in immunosuppression esp. in patients with cancer or transplant.  
 CC recipients.  
 CC Sequence 130 AA;

Query Match 100.0%; Score 15; DB 1; Length 130;  
 Best Local Similarity 25.0%; Pred. No. 4.03e+03;  
 Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 56 PDRFGSI 63  
 |  
 QY 2 PXXXXXXI 9

RESULT 11  
 ID R42675 standard; Protein; 167 AA.  
 AC R42675;  
 DT 21-APR-1994 (first entry)  
 DE Golden hamster Aphrodite precursor.  
 KW Aphrodite; pheromone; golden hamster; domestic animal breeding;  
 KW mammalian impotence; treatment.  
 OS Mesocricetus auratus.  
 FH Key Location/Qualifiers  
 FT peptide 1..16  
 FT /label= signal\_peptide  
 FT /note= "Claim 14"  
 FT protein 17..167  
 FT /label= mature\_Aphrodite

PN W09319173-A.  
 PD 30-SEP-1993.  
 PF 18-MAR-1993; E00621.  
 PR 18-MAR-1992; DE-208634.  
 PR 08-AUG-1992; DE-226340.  
 PR 07-NOV-1992; DE-237668.  
 PA (FORS) FORSMANN W.  
 PI Magert H; Maegert H;  
 DR WPI: 93-320740/40.  
 DR N-PSDB; Q49206.  
 PT DNA encoding aphrodite with an N-terminal signal peptide -  
 PT useful in the isolation of further aphrodite genes and as  
 PT peptide in medicaments for potency treatment  
 PS Claim 13; Page 46; 59pp; German.  
 CC This sequence represents the Golden hamster Aphrodite precursor.  
 CC The Aphrodite pheromone and its fragments are useful for treating  
 CC mammalian impotence, e.g. in domestic animal breeding programmes. A  
 CC closely similar Aphrodite sequence (differing only at one position  
 CC in the signal sequence) can be obtained from the Field hamster  
 CC (see R42674).  
 SQ Sequence 167 AA;

Query Match 100.0%; Score 15; DB 1; Length 167;  
 Best Local Similarity 25.0%; Pred. No. 4.03e+03;  
 Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 152 PVENILNI 159  
 |  
 QY 2 PXXXXXXI 9

RESULT 12  
 ID R15620 standard; protein; 174 AA.  
 AC R15620;  
 DT 25-MAR-1992 (first entry)  
 DE HBSAg pre-S region subtype adw.  
 KW T-cell epitope; vaccine; hepatitis B virus; antigen.  
 OS Synthetic.  
 PN W09117768-A.  
 PD 28-NOV-1991.  
 PF 10-MAY-1991; U03268.  
 PR 11-MAY-1990; US-522663.  
 PA SCRIPPS CLINIC & RE.

PI Millich DR, Thornton GB;  
 DR WPI: 91-369007/50.  
 PT Hepatitis B virus surface antigen epitope(s) - useful as vaccines,  
 PT immunogens or diagnostic reagents  
 PS Claim 1; Fig 1; 91pp; English.  
 CC The amino acid sequence is that of a pre-S T cell epitope polypeptide  
 CC of the pre-S (2) region of hepatitis B surface antigen d (HBSAg/d).  
 CC It can be used to prime or vaccinate a host to induce responsiveness  
 CC to HBV vaccine. The T cell epitope polypeptides can also be used as  
 CC immunogens that prime T cells that respond to native HBSAg B cell  
 CC epitope polypeptide. The T cell epitope polypeptides are also useful  
 CC as substitutes for carrier immunogens such as KtH and are safe,  
 CC defined and T cell-active. In addition to their use as vaccines, the  
 CC polypeptides can be used as immunogens for prodn. of antibodies. See  
 CC also R15617-R15622.  
 SQ Sequence 174 AA;

Query Match 100.0%; Score 15; DB 1; Length 174;  
 Best Local Similarity 25.0%; Pred. No. 4.03e+03;  
 Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 41 PDWDFNPI 48  
 |  
 QY 2 PXXXXXXI 9

RESULT 13  
 ID P30163 standard; peptide; 175 AA.  
 AC P30163;  
 DT 14-JUN-1992 (first entry)  
 DE Sequence encoded by a modified BamHI human interferon-alpha  
 DE gene fragment.  
 KW Yeast expression vector; Saccharomyces cerevisiae; promoter;  
 KW glycolytic enzyme; phosphoglycerate kinase.  
 OS Homo sapiens.  
 PN EP-73635-A.  
 PD 09-MAR-1983.  
 PF 17-AUG-1982; 408826.  
 PR 25-AUG-1981; GB-025934.  
 PR 23-MAR-1982; GB-008422.  
 PR 16-JUN-1982; GB-017496.  
 PA (KING/) KINGSMAN A J.  
 PA (CELL-) CELTECH LTD.  
 PI Kingsman S M; Kingsman A J.  
 DR WPI: 83-25386K/11.  
 DR N-PSDB; N30062.

PT Yeast expression vector for transforming yeasts - useful in  
 PT economic prodn. of polypeptide(s) esp. human interferon-alpha  
 PS Example; Fig 16; 45pp; English.  
 CC The inventors claim a yeast expression vector comprising a yeast  
 CC selective marker, a yeast replication origin and a yeast promoter  
 CC positioned relative to a unique restriction site. The yeast  
 CC promoter pref. comprises at least part of the 5' region of a gene  
 CC coding for glycolytic enzyme, esp. of the yeast PGK gene located up-  
 CC stream of the unique restriction site and at least part of the 3',  
 CC region of the PGK gene located downstream of the site. The vector  
 CC is used to express a polypeptide, eg. human interferon-alpha.  
 SQ Sequence 175 AA;

Query Match 100.0%; Score 15; DB 1; Length 175;  
 Best Local Similarity 25.0%; Pred. No. 4.03e+03;  
 Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 119 PLMKEDSI 126  
 |  
 QY 2 PXXXXXXI 9

RESULT 14  
 ID R30779 standard; Protein; 188 AA.  
 AC R30779;  
 DT 19-MAY-1993 (first entry)  
 DE p2130 polypeptide.

```
FT region 66..67 /label= loop_DE
FT 68..78 /label= beta-strand_E
FT 79..86 /label= loop_EF
FT 87..94 /label= beta-strand_F
FT 95..100 /label= loop_FG
FT 101..105 /label= beta-strand_G
FT misc_difference 46..51 /label= mutation
FT /note= "SRASGK -> AAAAGA"
PN WO9304173-A.
PD 04-MAR-1993.
PF 14-AUG-1992; U06860.
PR 14-AUG-1991; US-744768.
PR 07-MAY-1992; US-879495.
PA (GETH ) GENENTECH INC.
PI Jardieu PM, Presta LG;
PI WPI; 93-094004/11.
DR Polypeptide(s) binding to specific Fc epsilon receptors - act as
PT Ige antagonists; useful for treating and preventing Ige-mediated
PT disorders e.g. allergies
PS Disclosure; Page 73; 113pp; English.
CC Ige mutants were prep'd. to evaluate their effect on binding to
CC anti-IgE, esp. Maell, and to Fc epsilon RI and Fc epsilon RII.
CC Some of the mutants were designed to substitute for a specific
CC amino acid residue another residue with either similar or very
CC different charge or size.
CC Mutant 55 shows +ve/-ve binding to Fc epsilon RI and
CC +ve binding to Fc epsilon RII.
SQ Sequence 110 AA;

Query Match 100.0%; Score 15; DB 1; Length 110;
Best Local Similarity 25.0%; Pred. No. 4.03e+03;
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 14 PSPFDLFI 21
QY 2 PXXXXXXI 9

RESULT 9
ID R32026 standard; Protein; 110 AA.
AC R32026;
DT 05-JUL-1993 (first entry)
DE Variant Ige - mutant Emut 59.
KW High affinity; FCEH; low affinity; FCEL; Padlan;
KW Ige receptor; Fc; IgG1.
OS Homo sapiens.
FH Key Location/Qualifiers
FT region 7..12 /label= beta-strand_A
FT 13..24 /label= loop_AB
FT 25..33 /label= beta-strand_B
FT 34..42 /label= loop_BC
FT 43..48 /label= beta-strand_C
FT 49..57 /label= loop_CD
FT 58..65 /label= beta-strand_D
FT 66..67 /label= loop_DE
FT 68..78 /label= beta-strand_E
FT 79..86 /label= loop_EF
FT 87..94 /label= beta-strand_F
FT 95..100 /label= loop_FG
FT 101..105 /label= beta-strand_G
FT misc_difference 46..51 /label= mutation
FT /note= "SRASGK -> AAAAGA"
PN WO9304173-A.
PD 04-MAR-1993.
PF 14-AUG-1992; U06860.
PR 14-AUG-1991; US-744768.
PR 07-MAY-1992; US-879495.
PA (GETH ) GENENTECH INC.
PI Jardieu PM, Presta LG;
PI WPI; 93-094004/11.
DR Polypeptide(s) binding to specific Fc epsilon receptors - act as
PT Ige antagonists; useful for treating and preventing Ige-mediated
PT disorders e.g. allergies
PS Disclosure; Page 73; 113pp; English.
CC Ige mutants were prep'd. to evaluate their effect on binding to
CC anti-IgE, esp. Maell, and to Fc epsilon RI and Fc epsilon RII.
CC Some of the mutants were designed to substitute for a specific
CC amino acid residue another residue with either similar or very
CC different charge or size.
CC Mutant 55 shows +ve/-ve binding to Fc epsilon RI and
CC +ve binding to Fc epsilon RII.
SQ Sequence 110 AA;

Query Match 100.0%; Score 15; DB 1; Length 110;
Best Local Similarity 25.0%; Pred. No. 4.03e+03;
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 14 PSPFDLFI 21
QY 2 PXXXXXXI 9

RESULT 10
ID R23817 standard; Protein; 130 AA.
AC R23817;
DT 27-OCT-1992 (first entry)
DE Recombinant light chain variable domain (16).
KW Complementarity determining region; light chain variable domain;
KW antigen binding site; ligand; framework region; cancer; transplant.
OS Synthetic.
FH Key Location/Qualifiers
FT region 23..29 /label= CDR(e)
FT 45..53 /label= CDR(f)
FT 88..100 /label= CDR(d)
FT WO9206193-A.
PN 16-APR-1992.
PD 04-OCT-1991; G01726.
PR 05-OCT-1990; GB-021679.
PA (GORM/) GORMAN S D.
PI Gorman SD, Routledge EG, Waldmann H;
PI WPI; 92-150879/18.
DR Ligands and antibodies with binding affinity for CD3 antigen -
PT for treatment of immunosuppression e.g. in graft rejection, and
PT cancer, esp. lymphoid malignancies
PS Claim 7; Page 31; 49pp; English.
CC The sequence given is a recombinant human light chain variable
CC domain ligand containing the complementarity determining region
CC (CDR) given in R23736, R23737 and R23738. CDR's are found in the
CC variable domains of light and heavy chains which form the antigen
CC binding site, and act as connectors between the four framework regions.
CC It has been noted that there seem to be no characteristic features
CC which distinguish human from mouse or rat CDR's and they are
CC therefore immunologically identical. This ligand has binding affinity
```

Db		13 PERYRRI 20   2 PXXXXXI 9
Qy		
RESULT	7	
ID	R10523 standard; Protein; 85 AA.	
AC	R10523;	
DE	22-APR-1991 (first entry)	
DT	Non-A non-B hepatitis specific antigenic protein encoded by phage	
DE	clone lambda HC512.	
DE	Non-A non-B hepatitis; antigenic protein; NANBH; phage clone;	
KW	immunoassay; antibodies; diagnosis; ss.	
KW	Homo sapiens.	
OS	WO9101376-A.	
PN	PN	
PD	07-FEB-1991.	
PF	13-JUL-1990.	J00906
PR	14-JUL-1989; JP-182073.	
PR	19-JUL-1989; JP-184739.	
PR	22-JUL-1989; JP-189874.	
PR	27-JUL-1989; JP-192721.	
PR	29-JUL-1989; JP-195413.	
PR	03-AUG-1989; JP-200217.	
PR	10-AUG-1989; JP-205722.	
PR	21-SEP-1989; JP-243304.	
PR	22-SEP-1989; JP-245268.	
PR	19-OCT-1989; JP-270398.	
P1	(CHUS ) CHUGAI SEIYAKU KK.	
PI	Arima T, Yamamoto O, Tsuchiya M, Oshima M;	
DR	WPI; 91-058149/08.	
PPT	Antigenic protein specific for non-A, non-B hepatitis - and CDNA	
PPT	coding for it which corresponds to RNA of infected liver tissue	
PS	or serum	
PS	Disclosure; Fig 3: 69pp; Japanese.	
CCC	The DNA corresponds to an RNA isolated directly from infected human	
CCC	liver tissue or serum. The product may be used for the immunoassay	
CCC	of antibodies to NANBH antigen in samples of serum etc. for	
CCC	diagnostic purposes.	
CCC	See also Q10523-540.	
SQ	Sequence 85 AA;	
Query Match	100.0%; Score 15; DB 1; Length 85;	
Best Local Similarity	25.0%; Pred. No. 4.03e+03;	
Matches	2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;	
Db	13 PERYRRI 20   2 PXXXXXI 9	
Qy		
RESULT	8	
ID	R32025 standard; Protein; 110 AA.	
AC	R32025;	
DE	05-JUL-1993 (first entry)	
DT	Variant IGE - mutant Emut 55.	
DE	High affinity; FCEH; low affinity; FCEL; Padlan;	
KW	IgE receptor; Fc; IgG1.	
KW	Homo sapiens.	
OS	Homo sapiens.	
FFH	Key	Location/Qualifiers
FT	region	7..12
FT	/label= beta-strand_A	
FT	region	13..24
FT	/label= loop_AB	
FT	region	25..33
FT	/label= beta-strand_B	
FT	region	34..42
FT	/label= loop_BC	
FT	region	43..48
FT	/label= beta-strand_C	
FT	region	49..57
FT	/label= loop_CD	
FT	region	58..65
FT	/label= beta-strand_D	

KW amino-terminal fragment; diagnosis; screening; chemiluminescence assay.  
OS Synthetic.  
PN WO9321526-A.  
PD 28-OCT-1993.  
PF 03-MAR-1993; U01817.  
PR 15-APR-1992; US-868949.  
PR 26-OCT-1992; US-965971.  
PA (ATHE-) ATHENA NEUROSCIENCES INC.  
PA (ELIL) LILLY & CO ELI.  
PI Fritz LC, Schenk DB, Seubert PA;  
DR WPI: 93-351873/44.  
PT Monitoring beta amyloid precursor protein processing - involves  
PT detecting soluble fragments from cleavage at amino terminals of  
PT peptide, used to study Alzheimer's disease and potential drugs  
for it  
PS Disclosure; Page 10; 38pp; English.  
CC The peptide represents the C-terminal 25 residues of the beta-  
CC amyloid precursor protein amino terminal fragment. Such a peptide  
CC was used to raise antibodies which can be used in a chemiluminescence  
CC assay to probe human lumbar cerebrospinal fluid to detect Alzheimer's  
CC disease. Such an assay can be used to diagnose or monitor amyloid-  
CC related diseases in a patient and to screen and evaluate potential  
CC drugs for the treatment of these diseases.  
CC See also R42398-403.  
SQ Sequence 25 AA;

Query Match 100.0%; Score 15; DB 1; Length 25;  
Best Local Similarity 25.0%; Pred. No. 4.03e+03;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 8 PGSGLTNI 15  
| | |  
QY 2 PXXXXXXI 9

RESULT 3  
ID R41139 standard; peptide; 26 AA.  
AC R41139;  
DT 22-MAR-1994 (first entry)  
DE HCV peptide Xxf-2 (aa 393-416; E2/NS1 N-terminal).  
KW Human immunodeficiency virus; HIV; hepatitis C virus; HCV;  
KW non-A non-B hepatitis; NANBH; human T-cell lymphotropic virus; HTLV;  
KW epitope; antibody; blotting; diagnosis; detection; vaccine.  
OS Synthetic.  
FH Key  
FT Location/Qualifiers  
FT 1  
FT /note= "the N-terminal comprises (A)-(B)-(X)-Y; where  
FT B= biotin;  
FT X= biotinylation cpd. incorporated  
FT during synthesis;  
FT Y= bond or linking gp(s). which  
FT minimises steric hindrance,  
FT where Y is not a bond it is pref. 1-10  
FT residues of (same or different) glycine,  
FT beta-alanine, 4-aminobutyric acid,  
FT 5-aminovaleric acid or 6-aminohexanoic acid;  
FT parentthesis around B and X indicate opt. presence  
FT at the specified positions but B or X must be  
FT present in at least one of the positions shown,  
FT B interacts with the peptide to give a cpd.  
FT with greater diagnostic sensitivity;  
FT A (optional)= one or more amino acids, NH2 or  
FT gp. which modifies the N-terminus;  
FT Z= one or more amino acids, OH, NH2, or a  
FT linkage involving either of these 2 gps."  
FT modified\_site 26  
FT /note= "the C-terminal comprises Y-(X)-Z"

PN WO9318054-A.  
PD 16-SEP-1993.  
PF 08-MAR-1993; E00517.  
PR 06-MAR-1992; EP-400598.  
PA (INNO-) INNOGENETICS NV SA.  
PI De LEYS R;

DR WPI: 93-30397/38.  
PT New biotinylated peptide(s) corresp. to immuno-dominant  
PT epitope(s) - with increased antigenicity, useful in antibodies  
PT detection and vaccines against hepatitis C, HIV and HTLV  
PS Claim 4; Page 90-98; 133pp; English.  
CC Peptide compns. comprise at least one and pref. a combination of  
CC two, three, four or more biotinylated peptides chosen from the  
CC sequences given in R41058-R41166. The peptides represent  
CC immunologically important regions of viral proteins and are  
CC prepd. by solid phase peptide synthesis. The compns. are  
CC useful for the detection of antibodies to HCV, and/or HIV,  
CC and/or HTLV-I or II.  
SQ Sequence 26 AA;

Query Match 100.0%; Score 15; DB 1; Length 26;  
Best Local Similarity 25.0%; Pred. No. 4.03e+03;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 16 PKQNVHLI 23  
| | |  
QY 2 PXXXXXXI 9

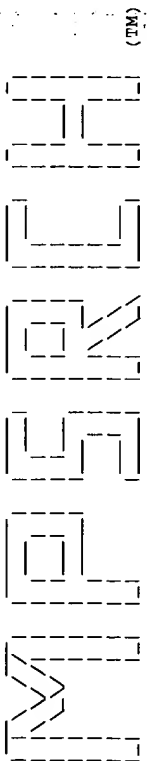
RESULT 4  
ID P80040 standard; protein; 42 AA.  
AC P80040;  
DT 12-NOV-1990 (first entry)  
DE Beta-human chorionic gonadotropin antigenically modified peptide (X).  
KW Human chorionic gonadotropin; isoimmunogen; thioester linkage;  
KW halo-acetic acid; antibodies.  
OS Synthetic.  
PN US4762913-A.  
PD 09-AUG-1988.  
PF 15-JUL-1987; 073769.  
PR 15-JUL-1987; US073769.  
PR 15-JUL-1987; US357892.  
PA (OHIS) OHIO STATE UNIV.  
PI Stevens VC;  
DR WPI: 88-242553/34.  
DT Antigenic modification of polypeptide(s) - by forming thioether linkage  
PT with carrier via active ester of halo-acetic acid.  
PS Claim 6; Page 95; 57pp; English.  
CC The peptide is chemically modified outside the body of an animal so  
CC that when injected into the animal they produce more antibodies against  
CC the unmodified protein than would injection of the unmodified protein or  
CC fragment alone. The chemical modification may be accomplished by  
CC attaching the peptide to carriers such as, e.g. bacterial toxoids, or by  
CC polymerisation of the peptide. The product can be administered to  
CC animals for the purpose of contraception, abortion or treatment of  
CC hormone-related disease states and disorders, hormone associated  
CC carcinomas, and to boost the animals resistance to exogenous proteins,  
CC e.g. viral.  
SQ Sequence 42 AA;

Query Match 100.0%; Score 15; DB 1; Length 42;  
Best Local Similarity 25.0%; Pred. No. 4.03e+03;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 25 PGPSDPTI 32  
| | |  
QY 2 PXXXXXXI 9

RESULT 5  
ID P81780 standard; protein; 51 AA.  
AC P81780;  
DT 10-MAR-1993 (revised)  
DT 15-NOV-1990 (first entry)  
DE Sequence encoded by open reading frame of cDNA corresponding to  
DE HIV-2 ROD genome  
KW LAV-II ROD; AIDS; Immunogen; antigen; vaccine; diagnostic.  
OS Human immunodeficiency virus ROD.  
PN WO8805440-A.

\*\*\*\*\*



\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPserch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Sat Apr 15 02:02:30 2000; MasPar time 3.07 Seconds  
Tabular output not generated. 69.466 Million cell updates/sec.

Title: >US-08-452-843-28  
Description: (1-9) from US08452843.pep  
Perfect Score: 15  
Sequence: 1 XPXXXXXXI 9

Scoring table: PAM 150  
Gap 15

Searched: 188963 seqs, 23686106 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: a-geneseq36  
1:geneseqp

Statistics: Mean 7.680; Variance 10.137; scale 0.758

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	15	100.0	16	R24939	HIV peptide ENV 312-327	4.03e+03
2	15	100.0	25	R42404	C-terminus of ATF-beta	4.03e+03
3	15	100.0	26	R41139	HCV peptide XH1-2 (aa	4.03e+03
4	15	100.0	42	P80040	Beta-human chorionic g	4.03e+03
5	15	100.0	51	P81780	Sequence encoded by op	4.03e+03
6	15	100.0	85	R10742	Non-A non-B hepatitis	4.03e+03
7	15	100.0	85	R10523	Non-A non-B hepatitis	4.03e+03
8	15	100.0	110	R32025	Variant Ige - mutant E	4.03e+03
9	15	100.0	110	R32026	Variant Ige - mutant E	4.03e+03
10	15	100.0	130	R23817	Recombinant light chai	4.03e+03
11	15	100.0	167	R42675	Golden hamster Aphrodi	4.03e+03
12	15	100.0	174	R15620	HSAg pre-S region sub	4.03e+03
13	15	100.0	175	P30163	Sequence encoded by a	4.03e+03
14	15	100.0	188	R30779	p130 polypeptide	4.03e+03
15	15	100.0	191	R07596	Animal somatotropin an	4.03e+03
16	15	100.0	206	R04127	Stem cell leukaemia (S	4.03e+03
17	15	100.0	212	R42428	GAG fusion protein wit	4.03e+03
18	15	100.0	214	R05857	Stem cell leukaemia (S	4.03e+03
19	15	100.0	249	R20573	Class B beta-lactamase	4.03e+03
20	15	100.0	266	R43576	Bovine adrenocorticotr	4.03e+03
21	15	100.0	270	R33043	Subtilisin 309 mutant	4.03e+03
22	15	100.0	270	R33042	Subtilisin 309 mutant	4.03e+03
23	15	100.0	286	R22688	Modified bla gene prod	4.03e+03

24	15	100.0	287	1	P80079	Sequence of 2,4-dichlo	4.03e+03
25	15	100.0	309	1	P94860	Expression plasmid pUC	4.03e+03
26	15	100.0	340	1	P94859	Expression plasmid pUC	4.03e+03
27	15	100.0	375	1	R48342	Rat stialyltransferase	4.03e+03
28	15	100.0	409	1	R53278	Ced-4 (W401X(sic)).	4.03e+03
29	15	100.0	439	1	R09258	t-PA variant d92-179,	4.03e+03
30	15	100.0	439	1	R09259	t-PA variant d92-179,	4.03e+03
31	15	100.0	459	1	R12550	Type I TNF receptor.	4.03e+03
32	15	100.0	459	1	P96202	Human muscarinic acety	4.03e+03
33	15	100.0	483	1	P93630	Sequence of rat transi	4.03e+03
34	15	100.0	476	1	R31023	Antibody D heavy chain	4.03e+03
35	15	100.0	482	1	R21109	NADH dehydrogenase sub	4.03e+03
36	15	100.0	503	1	R53285	Ced-3 (A449V).	4.03e+03
37	15	100.0	503	1	R45323	Ced-3 mutant I438.	4.03e+03
38	15	100.0	552	1	R11909	npRM-DNA able to form	4.03e+03
39	15	100.0	640	1	P81145	Sequence of fusion pro	4.03e+03
40	15	100.0	680	1	R23143	Mutant thermostable DN	4.03e+03
41	15	100.0	720	1	R15379	Pseudomonas SY77-glu	4.03e+03
42	15	100.0	720	1	R14445	Pseudomonas SY77-glu	4.03e+03
43	15	100.0	748	1	R24398	Prod. of the S gene of	4.03e+03
44	15	100.0	858	1	P81779	Sequence encoded by op	4.03e+03
45	15	100.0	924	1	R42380	Recombinant leukotoxin	4.03e+03

ALIGNMENTS

RESULT 1  
ID R24939 standard; Protein; 16 AA.  
AC R24939; 100.00 (first entry)  
DE HIV peptide ENV 312-327.  
KW Lipopeptide; lipoprotein; vaccine; cytotoxic T-cell; lymphocyte;  
KW HIV; human immunodeficiency virus; AIDS; cancer; tumour cells;  
KW CB1; CB2; CB3.  
OS Synthetic.  
PN EP-491628-A.  
PD 24-JUN-1992.  
PF 18-DEC-1991; 403446.  
PR 18-DEC-1990; FR-015870.  
PA (INSP ) INST PASTEUR LILLE.  
PA (INRM ) INERM INST NAT SANTE & RECH MEDICALE.  
PI Bouillon C, Gomard E, Gras-Masse H, Magne R,  
PI Martinon F, Sargheraert C, Tartar A, Levy JP;  
DR WPI; 92-209776/26.  
PT Lipopeptide(s) which stimulate cytotoxic T-cells - for treating  
PT HIV, parasitic infections and cancer  
PS Example; Page 18; 32pp; French.  
CC The sequence is that of peptide ENV 312-327 derived from the HIV,  
CC it is made by standard methods of solid phase peptide synthesis.  
CC It is used as part of lipoproteins CB1, CB2 and CB3 which comprise the  
CC peptide, and one or more chains derived from 10-20C fatty acids and/  
CC or modified sterol groups, these being coupled to alpha or epsilon  
CC amino groups of the peptide. The lipopeptides are useful in vaccines  
CC and acts by inducing cytotoxic T lymphocytes against the HIV  
CC virus antigen from which the peptide is derived.  
CC See also R24938 and R24940.  
SQ Sequence 16 AA;

Query Match 100.0%; Score 15; DB 1; Length 16;  
Best Local Similarity 25.0%; Pred. No. 4.03e+03;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 7 PGRAFTVI 14  
QY 2 PXXXXXXI 9

RESULT 2  
ID R42404 standard; peptide; 25 AA.  
AC R42404;  
DE 01-APR-1994 (first entry)  
DE C-terminus of ATF-beta APP.  
KW Amyloid precursor protein; Alzheimer's disease; fragments; inhibitors;



THIS PAGE BLANK (USPTO)

```
RP SEQUENCE FROM N.A.
RC STRAIN-DELTA H;
RX MEDLINE; 98037514.
RA SMITH D.R., DOUCETTE-STAMM L.A., DELOUGHERY C., LEE H.-M., DUBOIS J.,
RA ALDREGE T., BASHIRZADEH R., BLAKELY D., COOK R., GILBERT K.,
RA HARRISON D., HOANG L., KEAGLE P., LUMM W., POTHIER B., QIU D.,
RA SPADAFORA R., VICARE R., WANG Y., WIERZBOWSKI J., GIBSON R.,
RA JIWANI N., CARUSO A., BUSH D., SAFER H., PATWELL D., PRABHAKAR S.,
RA MCDUGALL S., SHIMER G., GOYAL A., PIETROVSKI S., CHURCH G.M.,
RA DANIELS C.J., MAO J.-I., RICE P., NOLLING J., REEVE J.N.;
RA "Complete genome sequence of Methanobacterium thermoautotrophicum
RT deltaH: functional analysis and comparative genomics.";
RL J. Bacteriol. 179:7135-7155(1997).
DR EMBL; AE000875; AAB8523.1; -.
KW Transferase.
SQ SEQUENCE 233 AA; 24618 MW; F88CE87C CRC32;

Query Match 100.0%; Score 15; DB 1; Length 233;
Best Local Similarity 25.0%; Pred. No. 4.94e+03;
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 225 PILPKRDV 232
Qy 2 PXXXXXXV 9

RESULT 15
ID O57741 PRELIMINARY; PRT; 238 AA.
AC O57741;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE 238AA LONG HYPOTHETICAL MODIFICATION METHYLASE.
GN PH0039.
OS Pyrococcus horikoshii.
OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=OT3;
RX MEDLINE; 98344137.
RA KAWARABAYASI Y., SAWADA M., HORIKAWA H., HAIKAWA Y., HINO Y.,
RA YAMAMOTO S., SEKINE M., BABA S., KOSUGI H., HOSOYAMA A., NAGAI Y.,
RA SAKAI M., OGURA K., OTUKA K., NAKAZAWA H., TAKAMIYA M., OHFUKU Y.,
RA FUNAHASHI T., TANAKA T., KUDOH Y., YAMAZAKI J., KUSHIDA N., OGUCHI A.,
RA AOKI K., NAKAMURA Y., ROBB T.F., HORIKOSHI K., MASUCHI Y., SHIZUYA H.,
RA KIKUCHI H.;
RA "Complete sequence and gene organization of the genome of a hyper-
RT thermophilic archaeobacterium, Pyrococcus horikoshii OT3.";
RL DNA Res. 5:55-76(1998).
DR EMBL; AP000001; BAA29107.1; -.
DR HSP; P20589; 1DCT.
DR PFAM; PF00145; DNA_methylase; 2.
KW Methyltransferase.
SQ SEQUENCE 238 AA; 28058 MW; 8AAF2532 CRC32;

Query Match 100.0%; Score 15; DB 1; Length 238;
Best Local Similarity 25.0%; Pred. No. 4.94e+03;
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 163 PYKPAPTV 170
Qy 2 PXXXXXXV 9
```

Search completed: Sat Apr 15 01:59:05 2000  
Job time : 90 secs.

FT NON\_TER 1 1  
 FT NON\_TER 193 193  
 SQ SEQUENCE 193 AA; 22215 MW; 97E67466 CRC32;  
 Query Match 100.0%; Score 15; DB 2; Length 193;  
 Best Local Similarity 25.0%; Pred. No. 4.94e+03;  
 Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
 Db 60 PSETIADV 67  
 QY 2 PXXXXXXV 9  
 RESULT 11  
 ID O07091 PRELIMINARY; PRT; 206 AA.  
 AC O07091;  
 DT 01-JUL-1997 (TREMBlrel. 04, Created)  
 DT 01-JUL-1997 (TREMBlrel. 04, Last sequence update)  
 DT 01-AUG-1998 (TREMBlrel. 07, Last annotation update)  
 DE CYTOCHROME C.  
 GN PSCC.  
 OS Chlorobium tepidum.  
 OC Bacteria; Green sulfur bacteria; Chlorobium.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA OH-OKA H., IWAKI M., ITOH S.;  
 RL Biochemistry 0:0-0(0).  
 DR EMBL; AB004460; BAA20402.1; -. 84BD2286 CRC32;  
 SQ SEQUENCE 206 AA; 22715 MW; 84BD2286 CRC32;

Query Match 100.0%; Score 15; DB 2; Length 206;  
 Best Local Similarity 25.0%; Pred. No. 4.94e+03;  
 Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
 Db 82 PLSIFIV 89  
 QY 2 PXXXXXXV 9  
 RESULT 12  
 ID O34947 PRELIMINARY; PRT; 210 AA.  
 AC O34947;  
 DT 01-JAN-1998 (TREMBlrel. 05, Created)  
 DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)  
 DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)  
 DE YOAZ.  
 GN YOAZ.  
 OS Bacillus subtilis.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
 OC Bacillus/Staphylococcus group; Bacillus.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA LAPIDUS A., GALLERON N., SOROKIN A., EHRLICH D.;  
 RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.  
 DT 2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RX MEDLINE; 98044033  
 RA KUNST F., OGASAWARA N., MOSTER I., ALBERTINI A.M., ALLONI G.,  
 RA AZEVEDO V., BERTERO M.G., BESSIERES P., BOLOTIN A., BORCHERT S.,  
 RA BORRIS R., BOURSIER L., BRANS A., BRAUN M., BRIGNELL S.C., BRON S.,  
 RA BROUILLET S., BRUSCHI C.V., CALDWELL B., CAPUANO V., CARTER N.M.,  
 RA CHOI S.K., CODANI J.J., CONNERTON I.F., CUMMINGS N.J., DANIEL R.A.,  
 RA DENIZOT F., DEVINE K.M., DUSTERHOFT A., EHRLICH S.D., EMERSON P.T.,  
 RA ENILAN K.D., ERRINGTON J., FABRET C., FERRARI E., FOULGER D.,  
 RA FRITZ C., FUJITA M., FUJITA Y., FUMA S., GALIZZI A., GALLERON N.,  
 RA GHIM S.Y., GLASER P., GOFFEAU A., GOLIGHTLY E.J., GRANDI G.,  
 RA GUISPIPI G., GUY B.J., HAGA K., HATECH J., HARWOOD C.R., HENAUT A.,  
 RA HILBERT H., HOLSAPPEL S., HOSONO S., HULLO M.F., ITAYA M., JONES L.,  
 RA JORIS B., KARAWATA D., KASAHARA Y., KLAERE-BLANCHARD M., KLEIN C.,  
 RA KOBAYASHI Y., KOETER P., KONINGSTEIN G., KROGH S., KUMANO M.,  
 RA KURITA K., LAPIDUS A., LARDINOIS S., LAUBER J., LAZAREVIC V.,  
 RA LEE S.M., LEVINE A., LIU H., MASUDA S., MAUEL C., MEDIGUE C.,

RA MEDINA N., MELLADO R.P., MIZUNO M., MOESTL D., NAKAI S., NOBACK M.,  
 RA NOONE D., O'REILLY M., OGAWA K., OGAWA A., OUDEGA B., PARK S.H.,  
 RA PARRO V., POHL T.M., PORTETELLE D., PORWOLLIK S., PRESCOTT A.M.,  
 RA PRESECAN E., PUJIC P., PURNELLE B., RAPOPORT G., REY M., REYNOLDS S.,  
 RA RIEGER M., RIVOLTA C., ROCHA E., ROCHE B., ROSE M., SADAIE Y.,  
 RA SATO T., SCANLAN E., SCHLEICH S., SCHROETER R., SCOFFONE F.,  
 RA SEKIGUCHI J., SEKOWSKA A., SEROR S.J., SERROR P., SHIN B.S., SOLDI B.,  
 RA SOROKIN A., TACCONI E., TAKAGI T., TAKAHASHI H., TAKEMARU K.,  
 RA TAKEUCHI M., TAKAKOSHI A., TANAKA T., TERPSTRA P., TOGNONI A.,  
 RA TOSATO V., UCHIYAMA S., VANDENBOL M., VANNIER F., VASSAROTTI A.,  
 RA VIARI A., WAMBUIT R., WEDLER E., WEDLER H., WEITZENEGGER T.,  
 RA WINTERS P., WIPAT A., YAMAMOTO H., YAMANE K., YASUMOTO K., YATA K.,  
 RA YOSHIDA K., YOSHIKAWA H.F., ZUMSTEIN E., YOSHIKAWA H., DANCHIN A.;  
 RT "The complete genome sequence of the gram-positive bacterium Bacillus  
 RT subtilis";  
 RL Nature 390:249-256(1997).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RA KUNST F., OGASAWARA N., YOSHIKAWA H., DANCHIN A.;  
 RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF027868; AAB84456.1; -.  
 DR EMBL; 299114; CAB13771.1; -.  
 SQ SEQUENCE 210 AA; 23420 MW; E0C4D02A CRC32;  
 Query Match 100.0%; Score 15; DB 2; Length 210;  
 Best Local Similarity 25.0%; Pred. No. 4.94e+03;  
 Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
 Db 167 PLEFAVEV 174  
 QY 2 PXXXXXXV 9  
 RESULT 13  
 ID O33795 PRELIMINARY; PRT; 231 AA.  
 AC O33795;  
 DT 01-JAN-1998 (TREMBlrel. 05, Created)  
 DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)  
 DT 01-AUG-1998 (TREMBlrel. 07, Last annotation update)  
 DE LAMBDA PHAGE L TAIL COMPONENT HOMOLOG.  
 OS Salmonella typhimurium.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Salmonella.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA DE GROOTE M., OCHSNER U.A., SHILOH M., NATHAN C., MCCORD J.M.,  
 RA DINAUER M.C., LIBBY S.J., VAZQUEZ-TORRES A., XU Y., FANG F.C.;  
 RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF007380; AAB62384.1; -.  
 SQ SEQUENCE 231 AA; 25460 MW; F52552A1 CRC32;  
 Query Match 100.0%; Score 15; DB 2; Length 231;  
 Best Local Similarity 25.0%; Pred. No. 4.94e+03;  
 Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
 Db 155 PRETDGVS 162  
 QY 2 PXXXXXXV 9  
 RESULT 14  
 ID O27106 PRELIMINARY; PRT; 233 AA.  
 AC O27106;  
 DT 01-JAN-1998 (TREMBlrel. 05, Created)  
 DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)  
 DT 01-AUG-1998 (TREMBlrel. 07, Last annotation update)  
 DE CDP-DIACYLGLYCEROL-SERINE O-PHOSPHATIDYLTRANSFERASE.  
 GN MTH1027.  
 OS Methanobacterium thermoautotrophicum.  
 OC Archaea; Euryarchaeota; Methanobacteriales; Methanobacteriaceae;  
 OC Methanobacterium.  
 RN [1]

```

AC O30467;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE ORF5 PROTEIN.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA GHIM S.-Y., JEONG Y.-M., CHOI S.-K., PARK S.-H.;
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF006665; AAB81147.1; -
DR HSP: P27830; LBKX.
SQ SEQUENCE 141 AA; 15982 MW; 9A812FC4 CRC32;

Query Match 100.0%; Score 15; DB 2; Length 141;
Best Local Similarity 25.0%; Pred. No. 4.94e+03;
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 11 PFFHNTV 18
QY 2 PXXXXXXV 9

RESULT 7
ID P71767 PRELIMINARY; PRT; 144 AA.
AC P71767;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-JAN-1998 (TrEMBLrel. 05, Last annotation update)
DE HYPOTHETICAL 15.2 KD PROTEIN CY277.08 PRECURSOR.
GN MRCY277.08.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RA MURPHY L., HARRIS D., BARRELL B.G., RAJANDREAM M.A., WALSH S.V.;
RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: SOME, TO DROSOPHILA AUBARIA GDH.
DR EMBL; Z79701; CAB02037.1; -
KW Hypothetical protein; Transmembrane; Signal.
FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 144 HYPOTHETICAL PROTEIN CY277.08.
FT TRANSMEM 24 44 POTENTIAL.
FT TRANSMEM 45 65 POTENTIAL.
SQ SEQUENCE 144 AA; 15159 MW; B5EBCD65 CRC32;

Query Match 100.0%; Score 15; DB 2; Length 144;
Best Local Similarity 25.0%; Pred. No. 4.94e+03;
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 46 PMWADGAV 53
QY 2 PXXXXXXV 9

RESULT 8
ID O30455 PRELIMINARY; PRT; 165 AA.
AC O30455;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE PARTICULATE METHANE MONOOXYGENASE PROTEIN A (FRAGMENT).
GN PMA.
OS unidentified eubacterium.
OC Bacteria; environmental samples.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 98174451.

RA McDONALD I.R., MURRELL J.C.;
RT "The particulate methane monooxygenase gene pmoA and its use as a
functional gene probe for methanotrophs.";
RL FEMS Microbiol. Lett. 156:205-210(1997).
DR EMBL; AF006047; AAC45950.1; -
KW Monooxygenase.
FT NON_TER 1 165
SQ SEQUENCE 165 AA; 18637 MW; 9D82C481 CRC32;

Query Match 100.0%; Score 15; DB 2; Length 165;
Best Local Similarity 25.0%; Pred. No. 4.94e+03;
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 35 PFGAVFAV 42
QY 2 PXXXXXXV 9

RESULT 9
ID O5YE17 PRELIMINARY; PRT; 187 AA.
AC O5YE17;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE 187AA LONG HYPOTHETICAL PROTEIN.
GN APE0753.
OS Aeropyrum pernix.
OC Archaea; Crenarchaeota; Aeropyrum.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K1;
RX MEDLINE; 99310339.
RA KAWARABAYASHI Y., HINO Y., HORIKAWA H., YAMAZAKI S., HAIKAWA Y.,
RA JIN-NO K., TAKAHASHI M., SEKINE M., BABA S., ANKAI A., KOSUGI H.,
RA HOSOMURA A., FUKUI S., NAGAI Y., NISHIJIMA K., NAKAZAWA H.,
RA TAKAYAMA M., MASUDA S., FUNAHASHI T., TANAKA T., KUDOH Y.,
RA YAMAZAKI J., KUSHIDA N., OGUCHI A., AOKI K., KUBOTA K., NAKAMURA Y.,
RA NOMURA N., SAKO Y., KIKUCHI H.;
RT "Complete genome sequence of an aerobic hyper-thermophilic
crenarchaeon, Aeropyrum pernix K1.";
RL DNA Res. 6:83-101(1999).
DR EMBL; AP000060; BAA79730.1; -
SQ SEQUENCE 187 AA; 20917 MW; 4C312EFF CRC32;

Query Match 100.0%; Score 15; DB 1; Length 187;
Best Local Similarity 25.0%; Pred. No. 4.94e+03;
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 59 PAFYPLQV 66
QY 2 PXXXXXXV 9

RESULT 10
ID O54607 PRELIMINARY; PRT; 193 AA.
AC O54607;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-JUN-1998 (TrEMBLrel. 06, Last annotation update)
DE POLYPHOSPHATE KINASE (FRAGMENT).
GN PPK.
OS Neisseria meningitidis.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
RN [1]
RP SEQUENCE FROM N.A.
RA MAIDEN M.C.J., BYGRAVES J.A., FEIL E., MORELLI G., RUSSELL J.E.,
RA URWIN R., ZHANG Q., ZHOU J., ZURTH K., CAUGANT D.A., FEAVERS I.M.,
RA ACHTYAN M., SPRATT B.G.;
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF037943; AAC08927.1; -
DR EMBL; AF037930; AAC08914.1; -
DR EMBL; AF037932; AAC08916.1; -

```

GN MJ1419.  
OS Methanococcus jannaschii.  
OC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;  
OC Methanococcus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 96337999.  
RA SUTTON G.G., WHITE O., OLSEN G.J., ZHOU L., FLEISCHMANN R.D.,  
RA BUTT C.J., BLAKE J.A., FITZGERALD L.M., CLAYTON R.A., GOCAYNE J.D.,  
RA KERAVAGE A.R., DOUGHERTY B.A., TOMB J.-F., ADAMS M.D., REICH C.I.,  
RA OVERBECK R., KIRKNESS E.F., WEINSTOCK K.G., MERRICK J.M., GLODEK A.,  
RA SCOTT J.L., GEOHAGAN S.N.M., WEIDMAN J.F., FUHRMANN J.L., NGUYEN D.,  
RA UTERBACK T.R., KELLEY J.M., PETERSON J.D., SADOW P.W., HANNA M.C.,  
RA COTTON M.D., ROBERTS K.M., HURST M.A., KAINE B.P., BORODOVSKY M.,  
RA KLENK H.-P., FRASER C.M., SMITH H.O., WOESE C.R., VENTER J.C.,  
RA "Complete genome sequence of the methanogenic archaeon, Methanococcus  
RT jannaschii".  
RL Science 273:1058-1073(1996).  
CC -!- SIMILARITY: TO M.THERMOAUTOTROPHICUM MTH1407 AND B.SUBTILIS YLOH  
CC PROTEIN.  
DR EMBL; U67582; AAB99429.1; -.  
DR TIGR: MJ1419; -.  
KW Hypothetical protein.  
SQ SEQUENCE 101 AA; 11222 MW; 4B33133E CRC32;

Query Match 100.0%; Score 15; DB 1; Length 101;  
Best Local Similarity 25.0%; Pred. No. 4.94e+03;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 20 PEEILKV 27  
|  
QY 2 PXXXXXXV 9

RESULT 3  
ID Q9YFC1 PRELIMINARY; PRT; 115 AA.  
AC Q9YFC1.  
DT 01-NOV-1999 (TrEMBLrel. 12, Created)  
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
DE 115AA LONG HYPOTHETICAL FERREDOXIN.  
GN APE0320.  
OS Aeropyrum pernix.  
OC Archaea; Crenarchaeota; Aeropyrum.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=K1.  
RX MEDLINE; 99310339.  
RA KARABAYASI Y., HINO Y., HORIKAWA H., YAMAZAKI S., HAIKAWA Y.,  
RA JIN-NO K., TAKAHASHI M., SEKINE M., BABA S., ANKAI A., KOSUGI H.,  
RA HOSOTAMA A., FUKUI S., NAGAI Y., NISHIJIMA K., NAKAZAWA H.,  
RA TAKAMIYA M., MASUDA S., FUNAHASHI T., TANAKA T., KUDOH Y.,  
RA YAMAZAKI J., KUSHIDA N., OGUCHI A., AOKI K., KUBOTA K., NAKAMURA Y.,  
RA NOMURA N., SAKO Y., KIKUCHI H.;  
RA "Complete genome sequence of an aerobic hyper-thermophilic  
RT crenarchaeon, Aeropyrum pernix K1.";  
RL DNA Res. 6:83-101(1999).  
DR EMBL; AP000059; BAA79275.1; -.  
DR PROSITE; PS00198; 4FE4S\_FERREDOXIN; 1.  
KW Iron-sulfur.  
SQ SEQUENCE 115 AA; 13387 MW; 2BDC676E CRC32;

Query Match 100.0%; Score 15; DB 1; Length 115;  
Best Local Similarity 25.0%; Pred. No. 4.94e+03;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 63 PVNCIKV 70  
|  
QY 2 PXXXXXXV 9

RESULT 4  
ID O57918 PRELIMINARY; PRT; 118 AA.

OS7918;  
AC 01-AUG-1998 (TrEMBLrel. 07, Created)  
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)  
DT 01-JAN-1999 (TrEMBLrel. 09, Last annotation update)  
DE 118AA LONG HYPOTHETICAL PROTEIN.  
GN PH0179.  
OS Pyrococcus horikoshii.  
OC Archaea; Euryarchaeota; Thermococcales; Pyrococcaceae; Pyrococcus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=OT3;  
RX MEDLINE; 98344137.  
RA KAWABAYASI Y., SAWADA M., HORIKAWA H., HAIKAWA Y., HINO Y.,  
RA YAMAMOTO S., SEKINE M., BABA S., KOSUGI H., HOSOTAMA A., NAGAI Y.,  
RA SAKAI M., OGURA K., OTUKA R., NAKAZAWA H., TAKAMIYA M., OHFUKU Y.,  
RA FUNAHASHI T., TANAKA T., KUDOH Y., YAMAZAKI J., KUSHIDA N., OGUCHI A.,  
RA AOKI K., NAKAMURA Y., ROBB T.F., HORIKOSHI K., MASUCHI Y., SHIZUYA H.,  
RA KIKUCHI H.;  
RA "Complete sequence and gene organization of the genome of a hyper-  
RT thermophilic archaeobacterium, Pyrococcus horikoshii OT3.";  
RL DNA Res. 5:55-76(1998).  
DR EMBL; AP000001; BAA29248.1; -.  
SQ SEQUENCE 118 AA; 12438 MW; 5C07DA6B CRC32;

Query Match 100.0%; Score 15; DB 1; Length 118;  
Best Local Similarity 25.0%; Pred. No. 4.94e+03;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 84 PKSPSPV 91  
|  
QY 2 PXXXXXXV 9

RESULT 5  
ID O27138 PRELIMINARY; PRT; 139 AA.  
AC O27138.  
DT 01-JAN-1998 (TrEMBLrel. 05, Created)  
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)  
DE 139AA LONG HYPOTHETICAL 15.6 KD PROTEIN.  
GN MTH1066.  
OS Methanobacterium thermoautotrophicum.  
OC Archaea; Euryarchaeota; Methanobacteriales; Methanobacteriaceae;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=DELTA H;  
RX MEDLINE; 98037514.  
RA SMITH D.R., DOUCETTE-STAMM L.A., DELOUGHERY C., LEE H.-M., DUBOIS J.,  
RA ALDREDGE T., BASHIRZADEH R., BLAKELY D., COOK R., GILBERT K.,  
RA HARRISON D., HOANG L., KEAGLE P., LUM W., POTHIER B., QIU D.,  
RA SPADAFORA R., VICARE R., WANG Y., WIERZBOWSKI J., GIBSON R.,  
RA JIWANI N., CARUSO A., BUSH D., SAFER H., PATWELL D., PRABHAKAR S.,  
RA MCDUGALL S., SHIMER G., GOYAL A., PIETROVSKI S., CHURCH G.M.,  
RA DANIELS C.J., MAO J.-I., RICE P., NOLLING J., REEVE J.N.;  
RA "Complete genome sequence of Methanobacterium thermoautotrophicum  
RT delta: functional analysis and comparative genomics.";  
RL J. Bacteriol. 179:7135-7155(1997).  
DR EMBL; AE000877; AAB85355.1; -.  
KW Hypothetical protein.  
SQ SEQUENCE 139 AA; 15587 MW; EE00CE2A CRC32;

Query Match 100.0%; Score 15; DB 1; Length 139;  
Best Local Similarity 25.0%; Pred. No. 4.94e+03;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 23 PFSKTEV 30  
|  
QY 2 PXXXXXXV 9

RESULT 6  
ID O30467 PRELIMINARY; PRT; 141 AA.

\*\*\*\*\*  
MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Sat Apr 15 01:57:35 2000; MasPar time 7.31 Seconds  
Tabular output not generated.  
Title: >US-08-452-843-27  
Description: (1-9) from US08452843.pep  
Perfect Score: 15  
Sequence: 1 XPXXXXXXV 9  
Scoring table: PAM 150  
Gap 15  
Searched: 225878 seqs, 69334122 residues  
Post-processing: Minimum Match 0%  
Listing first 45 summaries  
Database: sptrembl12  
1:sp\_archaea 2:sp\_bacteria 3:sp\_fungi 4:sp\_human  
5:sp\_invertebrate 6:sp\_mammal 7:sp\_mhc 8:sp\_organelle  
9:sp\_phage 10:sp\_plant 11:sp\_rodent 12:sp\_unclassified  
13:sp\_vertebrate 14:sp\_virus  
Statistics: Mean 10.140; Variance 6.383; scale 1.588  
Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.  
SUMMARIES  
Result No. Score Query Match Length DB ID Description Pred. No.  
1 15 100.0 95 2 P94797 10 KD CHAPERONIN (PROT 4.94e+03  
2 15 100.0 101 1 Q58814 HYPOTHETICAL PROTEIN M 4.94e+03  
3 15 100.0 115 1 O9YFC1 115AA LONG HYPOTHETICA 4.94e+03  
4 15 100.0 118 1 O57918 118AA LONG HYPOTHETICA 4.94e+03  
5 15 100.0 139 1 O27138 HYPOTHETICAL 15.6 KD P 4.94e+03  
6 15 100.0 141 2 O30467 ORF5 PROTEIN. 4.94e+03  
7 15 100.0 144 2 P71767 HYPOTHETICAL 15.2 KD P 4.94e+03  
8 15 100.0 165 2 O30455 PARTICULATE METHANE MO 4.94e+03  
9 15 100.0 187 1 O9Y817 187AA LONG HYPOTHETICA 4.94e+03  
10 15 100.0 193 2 O54607 POLYPHOSPHATE KINASE ( 4.94e+03  
11 15 100.0 206 2 O07091 CYTOCHROME C. 4.94e+03  
12 15 100.0 210 2 O34947 YOA2. 4.94e+03  
13 15 100.0 231 2 O33795 LAMBDA PHAGE L TAIL CO 4.94e+03  
14 15 100.0 233 1 O27106 CDP-DIACYLGLYCEROL-SER 4.94e+03  
15 15 100.0 238 1 O57741 238AA LONG HYPOTHETICA 4.94e+03  
16 15 100.0 256 2 O56146 CAMP FACTOR. 4.94e+03  
17 15 100.0 256 1 O58253 256AA LONG HYPOTHETICA 4.94e+03  
18 15 100.0 267 2 O52901 PEPTIDE SYNTHETASE (FR 4.94e+03  
19 15 100.0 300 2 O31348 ORF2 PROTEIN. 4.94e+03  
20 15 100.0 312 1 O27108 HYPOTHETICAL 34.0 KD P 4.94e+03

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Sat Apr 15 01:57:35 2000; MasPar time 7.31 Seconds  
Tabular output not generated.  
Title: >US-08-452-843-27  
Description: (1-9) from US08452843.pep  
Perfect Score: 15  
Sequence: 1 XPXXXXXXV 9  
Scoring table: PAM 150  
Gap 15  
Searched: 225878 seqs, 69334122 residues  
Post-processing: Minimum Match 0%  
Listing first 45 summaries  
Database: sptrembl12  
1:sp\_archaea 2:sp\_bacteria 3:sp\_fungi 4:sp\_human  
5:sp\_invertebrate 6:sp\_mammal 7:sp\_mhc 8:sp\_organelle  
9:sp\_phage 10:sp\_plant 11:sp\_rodent 12:sp\_unclassified  
13:sp\_vertebrate 14:sp\_virus  
Statistics: Mean 10.140; Variance 6.383; scale 1.588  
Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.  
SUMMARIES  
Result No. Score Query Match Length DB ID Description Pred. No.  
1 15 100.0 95 2 P94797 10 KD CHAPERONIN (PROT 4.94e+03  
2 15 100.0 101 1 Q58814 HYPOTHETICAL PROTEIN M 4.94e+03  
3 15 100.0 115 1 O9YFC1 115AA LONG HYPOTHETICA 4.94e+03  
4 15 100.0 118 1 O57918 118AA LONG HYPOTHETICA 4.94e+03  
5 15 100.0 139 1 O27138 HYPOTHETICAL 15.6 KD P 4.94e+03  
6 15 100.0 141 2 O30467 ORF5 PROTEIN. 4.94e+03  
7 15 100.0 144 2 P71767 HYPOTHETICAL 15.2 KD P 4.94e+03  
8 15 100.0 165 2 O30455 PARTICULATE METHANE MO 4.94e+03  
9 15 100.0 187 1 O9Y817 187AA LONG HYPOTHETICA 4.94e+03  
10 15 100.0 193 2 O54607 POLYPHOSPHATE KINASE ( 4.94e+03  
11 15 100.0 206 2 O07091 CYTOCHROME C. 4.94e+03  
12 15 100.0 210 2 O34947 YOA2. 4.94e+03  
13 15 100.0 231 2 O33795 LAMBDA PHAGE L TAIL CO 4.94e+03  
14 15 100.0 233 1 O27106 CDP-DIACYLGLYCEROL-SER 4.94e+03  
15 15 100.0 238 1 O57741 238AA LONG HYPOTHETICA 4.94e+03  
16 15 100.0 256 2 O56146 CAMP FACTOR. 4.94e+03  
17 15 100.0 256 1 O58253 256AA LONG HYPOTHETICA 4.94e+03  
18 15 100.0 267 2 O52901 PEPTIDE SYNTHETASE (FR 4.94e+03  
19 15 100.0 300 2 O31348 ORF2 PROTEIN. 4.94e+03  
20 15 100.0 312 1 O27108 HYPOTHETICAL 34.0 KD P 4.94e+03

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Sat Apr 15 01:57:35 2000; MasPar time 7.31 Seconds  
Tabular output not generated.  
Title: >US-08-452-843-27  
Description: (1-9) from US08452843.pep  
Perfect Score: 15  
Sequence: 1 XPXXXXXXV 9  
Scoring table: PAM 150  
Gap 15  
Searched: 225878 seqs, 69334122 residues  
Post-processing: Minimum Match 0%  
Listing first 45 summaries  
Database: sptrembl12  
1:sp\_archaea 2:sp\_bacteria 3:sp\_fungi 4:sp\_human  
5:sp\_invertebrate 6:sp\_mammal 7:sp\_mhc 8:sp\_organelle  
9:sp\_phage 10:sp\_plant 11:sp\_rodent 12:sp\_unclassified  
13:sp\_vertebrate 14:sp\_virus  
Statistics: Mean 10.140; Variance 6.383; scale 1.588  
Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.  
SUMMARIES  
Result No. Score Query Match Length DB ID Description Pred. No.  
1 15 100.0 95 2 P94797 10 KD CHAPERONIN (PROT 4.94e+03  
2 15 100.0 101 1 Q58814 HYPOTHETICAL PROTEIN M 4.94e+03  
3 15 100.0 115 1 O9YFC1 115AA LONG HYPOTHETICA 4.94e+03  
4 15 100.0 118 1 O57918 118AA LONG HYPOTHETICA 4.94e+03  
5 15 100.0 139 1 O27138 HYPOTHETICAL 15.6 KD P 4.94e+03  
6 15 100.0 141 2 O30467 ORF5 PROTEIN. 4.94e+03  
7 15 100.0 144 2 P71767 HYPOTHETICAL 15.2 KD P 4.94e+03  
8 15 100.0 165 2 O30455 PARTICULATE METHANE MO 4.94e+03  
9 15 100.0 187 1 O9Y817 187AA LONG HYPOTHETICA 4.94e+03  
10 15 100.0 193 2 O54607 POLYPHOSPHATE KINASE ( 4.94e+03  
11 15 100.0 206 2 O07091 CYTOCHROME C. 4.94e+03  
12 15 100.0 210 2 O34947 YOA2. 4.94e+03  
13 15 100.0 231 2 O33795 LAMBDA PHAGE L TAIL CO 4.94e+03  
14 15 100.0 233 1 O27106 CDP-DIACYLGLYCEROL-SER 4.94e+03  
15 15 100.0 238 1 O57741 238AA LONG HYPOTHETICA 4.94e+03  
16 15 100.0 256 2 O56146 CAMP FACTOR. 4.94e+03  
17 15 100.0 256 1 O58253 256AA LONG HYPOTHETICA 4.94e+03  
18 15 100.0 267 2 O52901 PEPTIDE SYNTHETASE (FR 4.94e+03  
19 15 100.0 300 2 O31348 ORF2 PROTEIN. 4.94e+03  
20 15 100.0 312 1 O27108 HYPOTHETICAL 34.0 KD P 4.94e+03

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Sat Apr 15 01:57:35 2000; MasPar time 7.31 Seconds  
Tabular output not generated.  
Title: >US-08-452-843-27  
Description: (1-9) from US08452843.pep  
Perfect Score: 15  
Sequence: 1 XPXXXXXXV 9  
Scoring table: PAM 150  
Gap 15  
Searched: 225878 seqs, 69334122 residues  
Post-processing: Minimum Match 0%  
Listing first 45 summaries  
Database: sptrembl12  
1:sp\_archaea 2:sp\_bacteria 3:sp\_fungi 4:sp\_human  
5:sp\_invertebrate 6:sp\_mammal 7:sp\_mhc 8:sp\_organelle  
9:sp\_phage 10:sp\_plant 11:sp\_rodent 12:sp\_unclassified  
13:sp\_vertebrate 14:sp\_virus  
Statistics: Mean 10.140; Variance 6.383; scale 1.588  
Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.  
SUMMARIES  
Result No. Score Query Match Length DB ID Description Pred. No.  
1 15 100.0 95 2 P94797 10 KD CHAPERONIN (PROT 4.94e+03  
2 15 100.0 101 1 Q58814 HYPOTHETICAL PROTEIN M 4.94e+03  
3 15 100.0 115 1 O9YFC1 115AA LONG HYPOTHETICA 4.94e+03  
4 15 100.0 118 1 O57918 118AA LONG HYPOTHETICA 4.94e+03  
5 15 100.0 139 1 O27138 HYPOTHETICAL 15.6 KD P 4.94e+03  
6 15 100.0 141 2 O30467 ORF5 PROTEIN. 4.94e+03  
7 15 100.0 144 2 P71767 HYPOTHETICAL 15.2 KD P 4.94e+03  
8 15 100.0 165 2 O30455 PARTICULATE METHANE MO 4.94e+03  
9 15 100.0 187 1 O9Y817 187AA LONG HYPOTHETICA 4.94e+03  
10 15 100.0 193 2 O54607 POLYPHOSPHATE KINASE ( 4.94e+03  
11 15 100.0 206 2 O07091 CYTOCHROME C. 4.94e+03  
12 15 100.0 210 2 O34947 YOA2. 4.94e+03  
13 15 100.0 231 2 O33795 LAMBDA PHAGE L TAIL CO 4.94e+03  
14 15 100.0 233 1 O27106 CDP-DIACYLGLYCEROL-SER 4.94e+03  
15 15 100.0 238 1 O57741 238AA LONG HYPOTHETICA 4.94e+03  
16 15 100.0 256 2 O56146 CAMP FACTOR. 4.94e+03  
17 15 100.0 256 1 O58253 256AA LONG HYPOTHETICA 4.94e+03  
18 15 100.0 267 2 O52901 PEPTIDE SYNTHETASE (FR 4.94e+03  
19 15 100.0 300 2 O31348 ORF2 PROTEIN. 4.94e+03  
20 15 100.0 312 1 O27108 HYPOTHETICAL 34.0 KD P 4.94e+03

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Sat Apr 15 01:57:35 2000; MasPar time 7.31 Seconds  
Tabular output not generated.  
Title: >US-08-452-843-27  
Description: (1-9) from US08452843.pep  
Perfect Score: 15  
Sequence: 1 XPXXXXXXV 9  
Scoring table: PAM 150  
Gap 15  
Searched: 225878 seqs, 69334122 residues  
Post-processing: Minimum Match 0%  
Listing first 45 summaries  
Database: sptrembl12  
1:sp\_archaea 2:sp\_bacteria 3:sp\_fungi 4:sp\_human  
5:sp\_invertebrate 6:sp\_mammal 7:sp\_mhc 8:sp\_organelle  
9:sp\_phage 10:sp\_plant 11:sp\_rodent 12:sp\_unclassified  
13:sp\_vertebrate 14:sp\_virus  
Statistics: Mean 10.140; Variance 6.383; scale 1.588  
Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.  
SUMMARIES  
Result No. Score Query Match Length DB ID Description Pred. No.  
1 15 100.0 95 2 P94797 10 KD CHAPERONIN (PROT 4.94e+03  
2 15 100.0 101 1 Q58814 HYPOTHETICAL PROTEIN M 4.94e+03  
3 15 100.0 115 1 O9YFC1 115AA LONG HYPOTHETICA 4.94e+03  
4 15 100.0 118 1 O57918 118AA LONG HYPOTHETICA 4.94e+03  
5 15 100.0 139 1 O27138 HYPOTHETICAL 15.6 KD P 4.94e+03  
6 15 100.0 141 2 O30467 ORF5 PROTEIN. 4.94e+03  
7 15 100.0 144 2 P71767 HYPOTHETICAL 15.2 KD P 4.94e+03  
8 15 100.0 165 2 O30455 PARTICULATE METHANE MO 4.94e+03  
9 15 100.0 187 1 O9Y817 187AA LONG HYPOTHETICA 4.94e+03  
10 15 100.0 193 2 O54607 POLYPHOSPHATE KINASE ( 4.94e+03  
11 15 100.0 206 2 O07091 CYTOCHROME C. 4.94e+03  
12 15 100.0 210 2 O34947 YOA2. 4.94e+03  
13 15 100.0 231 2 O33795 LAMBDA PHAGE L TAIL CO 4.94e+03  
14 15 100.0 233 1 O27106 CDP-DIACYLGLYCEROL-SER 4.94e+03  
15 15 100.0 238 1 O57741 238AA LONG HYPOTHETICA 4.94e+03  
16 15 100.0 256 2 O56146 CAMP FACTOR. 4.94e+03  
17 15 100.0 256 1 O58253 256AA LONG HYPOTHETICA 4.94e+03  
18 15 100.0 267 2 O52901 PEPTIDE SYNTHETASE (FR 4.94e+03  
19 15 100.0 300 2 O31348 ORF2 PROTEIN. 4.94e+03  
20 15 100.0 312 1 O27108 HYPOTHETICAL 34.0 KD P 4.94e+03

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Sat Apr 15 01:57:35 2000; MasPar time 7.31 Seconds  
Tabular output not generated.  
Title: >US-08-452-843-27  
Description: (1-9) from US08452843.pep  
Perfect Score: 15  
Sequence: 1 XPXXXXXXV 9  
Scoring table: PAM 150  
Gap 15  
Searched: 225878 seqs, 69334122 residues  
Post-processing: Minimum Match 0%  
Listing first 45 summaries  
Database: sptrembl12  
1:sp\_archaea 2:sp\_bacteria 3:sp\_fungi 4:sp\_human  
5:sp\_invertebrate 6:sp\_mammal 7:sp\_mhc 8:sp\_organelle  
9:sp\_phage 10:sp\_plant 11:sp\_rodent 12:sp\_unclassified  
13:sp\_vertebrate 14:sp\_virus  
Statistics: Mean 10.140; Variance 6.383; scale 1.588  
Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.  
SUMMARIES  
Result No. Score Query Match Length DB ID Description Pred. No.  
1 15 100.0 95 2 P94797 10 KD CHAPERONIN (PROT 4.94e+03  
2 15 100.0 101 1 Q58814 HYPOTHETICAL PROTEIN M 4.94e+03  
3 15 100.0 115 1 O9YFC1 115AA LONG HYPOTHETICA 4.94e+03  
4 15 100.0 118 1 O57918 118AA LONG HYPOTHETICA 4.94e+03  
5 15 100.0 139 1 O27138 HYPOTHETICAL 15.6 KD P 4.94e+03  
6 15 100.0 141 2 O30467 ORF5 PROTEIN. 4.94e+03  
7 15 100.0 144 2 P71767 HYPOTHETICAL 15.2 KD P 4.94e+03  
8 15 100.0 165 2 O30455 PARTICULATE METHANE MO 4.94e+03  
9 15 100.0 187 1 O9Y817 187AA LONG HYPOTHETICA 4.94e+03  
10 15 100.0 193 2 O54607 POLYPHOSPHATE KINASE ( 4.94e+03  
11 15 100.0 206 2 O07091 CYTOCHROME C. 4.94e+03  
12 15 100.0 210 2 O34947 YOA2. 4.94e+03  
13 15 100.0 231 2 O33795 LAMBDA PHAGE L TAIL CO 4.94e+03  
14 15 100.0 233 1 O27106 CDP-DIACYLGLYCEROL-SER 4.94e+03  
15 15 100.0 238 1 O57741 238AA LONG HYPOTHETICA 4.94e+03  
16 15 100.0 256 2 O56146 CAMP FACTOR. 4.94e+03  
17 15 100.0 256 1 O58253 256AA LONG HYPOTHETICA 4.94e+03  
18 15 100.0 267 2 O52901 PEPTIDE SYNTHETASE (FR 4.94e+03  
19 15 100.0 300 2 O31348 ORF2 PROTEIN. 4.94e+03  
20 15 100.0 312 1 O27108 HYPOTHETICAL 34.0 KD P 4.94e+03

21 15 100.0 313 2 O34920 TRANSCRIPTION REGULATO 4.94e+03  
22 15 100.0 315 1 O50082 315AA LONG HYPOTHETICA 4.94e+03  
23 15 100.0 339 2 P71763 HYPOTHETICAL 38.3 KD P 4.94e+03  
24 15 100.0 339 1 O9YCC1 339AA LONG HYPOTHETICA 4.94e+03  
25 15 100.0 341 1 O9YDV4 341AA LONG HYPOTHETICA 4.94e+03  
26 15 100.0 344 2 Q48025 OUTER MEMBRANE PROTEIN 4.94e+03  
27 15 100.0 344 2 Q48024 CORE PROTEIN (FRAGMENT 4.94e+03  
28 15 100.0 350 2 O52663 352AA LONG HYPOTHETICA 4.94e+03  
29 15 100.0 352 1 O58250 379AA LONG HYPOTHETICA 4.94e+03  
30 15 100.0 379 1 O9YEW4 379AA LONG HYPOTHETICA 4.94e+03  
31 15 100.0 382 2 O58353 ALKALINE SERINE PROTEA 4.94e+03  
32 15 100.0 402 1 O29593 CONSERVED HYPOTHETICAL 4.94e+03  
33 15 100.0 417 2 O50201 HYALURONAN SYNTHASE. 4.94e+03  
34 15 100.0 430 1 O27142 CONSERVED PROTEIN. 4.94e+03  
35 15 100.0 438 1 O9YEB2 438AA LONG HYPOTHETICA 4.94e+03  
36 15 100.0 440 1 O59178 440AA LONG HYPOTHETICA 4.94e+03  
37 15 100.0 441 1 O59179 441AA LONG HYPOTHETICA 4.94e+03  
38 15 100.0 443 1 Q58516 HYPOTHETICAL PROTEIN M 4.94e+03  
39 15 100.0 469 1 O9YFC4 469AA LONG HYPOTHETICA 4.94e+03  
40 15 100.0 478 1 O9YEC7 478AA LONG HYPOTHETICA 4.94e+03  
41 15 100.0 538 1 O9YCC0 538AA LONG HYPOTHETICA 4.94e+03  
42 15 100.0 547 1 O28848 PHENYLALANYL-TRNA SYNT 4.94e+03  
43 15 100.0 718 1 O58791 718AA LONG HYPOTHETICA 4.94e+03  
44 15 100.0 1097 2 P72196 TONB-LINKED ADHESIN PR 4.94e+03  
45 15 100.0 1556 2 Q07270 SR PROTEIN. 4.94e+03

ALIGNMENTS

RESULT 1  
ID P94797 PRELIMINARY; PRT; 95 AA.  
AC P94797;  
DT 01-MAY-1997 (Tremblrel. 03, Created)  
DT 01-MAY-1997 (Tremblrel. 03, Last sequence update)  
DE 01-NOV-1999 (Tremblrel. 12, Last annotation update)  
DE 10 KD CHAPERONIN (PROTEIN CPN10) (PROTEIN GROES).  
GN GROES.  
OS Francisella tularensis.  
OC Bacteria; Proteobacteria; gamma subdivision; Thiomicrospira group;  
OC Francisella group; Francisella.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-LVS;  
RA ERICSSON M., GOLOVLIQV I., SJSTEDT A., TRNVIK A.;  
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: BINDS TO CPN60 IN THE PRESENCE OF MG-ATP AND SUPPRESS  
CC THE ATPASE ACTIVITY OF THE LATER (BY SIMILARITY).  
CC -!- SUBUNIT: HEPTAMER OF 7 SUBUNITS ARRANGED IN A RING (BY  
CC SIMILARITY).  
CC -!- SIMILARITY: BELONGS TO THE GROES CHAPERONIN FAMILY.  
DR EMBL: X98853; CAA67359.1; -.  
DR PROSITE: PS00681; CHAPERONINS\_CPN10; 1.  
DR PFWA; PFW0166; cpn10; 1.  
DR PRINTS; PR00297; CHAPERONIN10.  
KW Chaperone.  
SQ SEQUENCE 95 AA; 10272 MW; 96AGB781 CRC32;  
Query Match 100.0%; Score 15; DB 2; Length 95;  
Best Local Similarity 25.0%; Pred. No. 4.94e+03;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 5 PLQDRVLV 12  
QY 2 PXXXXXXV 9  
RESULT 2  
ID Q58814 PRELIMINARY; PRT; 101 AA.  
AC Q58814;  
DT 01-JUN-1998 (Tremblrel. 06, Created)  
DT 01-JUN-1998 (Tremblrel. 06, Last sequence update)  
DT 01-JUN-1998 (Tremblrel. 06, Last annotation update)  
DE HYPOTHETICAL PROTEIN MJ1419.

THIS PAGE BLANK

```

CC      or send an email to license@lsb-sib.ch).
CC      -----
DR      EMBL: X97499; CAA66130.1; -.
DR      PROSITE; PS00901; CYS_SYNTHASE; 1.
DR      PFAM; PF00291; S_T_dehydratase; 1.
KW      Lyase; Cysteine biosynthesis; Pyridoxal phosphate.
FT      BINDING      43      43      PYRIDOXAL PHOSPHATE (BY SIMILARITY).
SQ      SEQUENCE      339 AA; 37053 MW; F6B32E5E CRC32;

Query Match      100.0%; Score 15; DB 1; Length 339;
Best Local Similarity 25.0%; Pred. No. 4.56e+03;
Matches      2; Conservative      0; Mismatches      6; Indels      0; Gaps      0;

Db      198 PSMRVGV 205
Qy      |
        2 PXXXXXXV 9

Search completed: Sat Apr 15 01:57:18 2000
Job time : 39 secs.

```





CC Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;  
CC Rhodobacter.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA BURKE D.H., ALBERTI M., ARMSTRONG G.A., HEARST J.E.;  
RL Submitted (NOV-1991) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP PRELIMINARY SEQUENCE FROM N.A.  
RC STRAIN=SB1003, AND BEC404;  
RX MEDLINE: 89313663.  
RA ARMSTRONG G.A., ALBERTI M., LEACH F., HEARST J.E.;  
RL "Nucleotide sequence, organization, and nature of the protein  
RT products of the carotenoid biosynthesis gene cluster of Rhodobacter  
RT capsulatus";  
RL Mol. Gen. Genet. 216:254-268(1989).  
CC -|- PATHWAY: CAROTENOID AND CHLOROPHYLL BIOSYNTHESIS.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
DR EMBL; Z11165; CAA7539.1; -;  
DR EMBL; X52291; CAA36532.1; -;  
DR PIR; S04401; S04401.  
DR PIR; S17822; S17822.  
KW Photosynthesis; Chlorophyll biosynthesis; Carotenoid biosynthesis;  
KW Oxidoreductase.  
SQ SEQUENCE 241 AA; 27004 MW; 59085F33 CRC32;  
  
Query Match 100.0%; Score 15; DB 1; Length 241;  
Best Local Similarity 25.0%; Pred. No. 4.56e+03;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
  
Db 30 PLNDEPRV 37  
|  
|  
Qy 2 PXXXXXXV 9  
  
RESULT 11  
ID ADH\_DRODI STANDARD; PRT; 253 AA.  
AC P22245;  
DT 01-AUG-1991 (Rel. 19, Created)  
DT 01-MAY-1992 (Rel. 22, Last sequence update)  
DT 01-OCT-1996 (Rel. 34, Last annotation update)  
DE ALCOHOL DEHYDROGENASE (EC 1.1.1.1).  
GN ADH.  
OS Drosophila differens (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 91163323.  
RA ROWAN R.G., HUNT J.A.;  
RT "Rates of DNA change and phylogeny from the DNA sequences of the  
RT alcohol dehydrogenase gene for five closely related species of  
RT Hawaiian Drosophila";  
RL Mol. Biol. Evol. 8:49-70(1991).  
CC -|- CATALYTIC ACTIVITY: ALCOHOL + NAD(+) - ALDEHYDE OR KETONE + NADH.  
CC -|- SUBUNIT: HOMODIMER.  
CC -|- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES  
CC FAMILY (SDR).  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
DR EMBL; M63303; AAA28350.1; -;  
DR FLYBASE; FBgn0012249; Ddif\Adh.  
DR PROSITE; PS00061; ADH\_SHORT; 1.  
DR PFAM; PF00106; adh\_short; 1.  
DR PFAM; PF00663; adh\_short\_C; 1.  
KW Oxidoreductase; NAD.  
FT INIT\_MET 0 0 BY SIMILARITY.  
FT NP\_BIND 9 32 NAD (BY SIMILARITY).  
FT ACT\_SITE 150 150 BY SIMILARITY.  
SQ SEQUENCE 253 AA; 27446 MW; 0B671F3F CRC32;  
  
Query Match 100.0%; Score 15; DB 1; Length 253;  
Best Local Similarity 25.0%; Pred. No. 4.56e+03;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
  
Db 60 PYDVTVPV 67  
|  
|  
Qy 2 PXXXXXXV 9

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).  
CC  
DR EMBL; M63303; AAA28350.1; -;  
DR FLYBASE; FBgn0012249; Ddif\Adh.  
DR PROSITE; PS00061; ADH\_SHORT; 1.  
DR PFAM; PF00106; adh\_short; 1.  
DR PFAM; PF00663; adh\_short\_C; 1.  
KW Oxidoreductase; NAD.  
FT INIT\_MET 0 0 BY SIMILARITY.  
FT NP\_BIND 9 32 NAD (BY SIMILARITY).  
FT ACT\_SITE 150 150 BY SIMILARITY.  
SQ SEQUENCE 253 AA; 27320 MW; B220785F CRC32;  
  
Query Match 100.0%; Score 15; DB 1; Length 253;  
Best Local Similarity 25.0%; Pred. No. 4.56e+03;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
  
Db 60 PYDVTVPV 67  
|  
|  
Qy 2 PXXXXXXV 9  
  
RESULT 12  
ID ADH\_DROFL STANDARD; PRT; 253 AA.  
AC P48585;  
DT 01-FEB-1996 (Rel. 33, Created)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE ALCOHOL DEHYDROGENASE (EC 1.1.1.1).  
GN ADH.  
OS Drosophila flavomontana (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 96139062.  
RA NURMINSKY D.I., MORIYAMA E.N., LOZOVSKAYA E.R., HARTL D.L.;  
RT "Molecular phylogeny and genome evolution in the Drosophila virilis  
RT species group: duplications of the alcohol dehydrogenase gene";  
RL Mol. Biol. Evol. 13:132-149(1996).  
CC -|- CATALYTIC ACTIVITY: ALCOHOL + NAD(+) - ALDEHYDE OR KETONE + NADH.  
CC -|- SUBUNIT: HOMODIMER.  
CC -|- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES  
CC FAMILY (SDR).  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
DR EMBL; U26838; BAB02624.1; -;  
DR FLYBASE; FBgn0013805; Dfla\Adh.  
DR PROSITE; PS00061; ADH\_SHORT; 1.  
DR PFAM; PF00106; adh\_short; 1.  
DR PFAM; PF00663; adh\_short\_C; 1.  
KW Oxidoreductase; NAD.  
FT INIT\_MET 0 0 BY SIMILARITY.  
FT NP\_BIND 9 32 NAD (BY SIMILARITY).  
FT ACT\_SITE 150 150 BY SIMILARITY.  
SQ SEQUENCE 253 AA; 27446 MW; 0B671F3F CRC32;  
  
Query Match 100.0%; Score 15; DB 1; Length 253;  
Best Local Similarity 25.0%; Pred. No. 4.56e+03;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
  
Db 60 PYDVTVPV 67  
|  
|  
Qy 2 PXXXXXXV 9



SQ SEQUENCE 134 AA; 15218 MW; BF2B535D CRC32;

Query Match 100.0%; Score 15; DB 1; Length 134;

Best Local Similarity 25.0%; Pred. No. 4.56e+03; Mismatches 6; Indels 0; Gaps 0;

Db 16 PQDVFV 23

QY 2 PXXXXXXV 9

RESULT 5

ID D3\_ONCVO STANDARD; PRT; 134 AA.

AC P54188;

DT 01-OCT-1996 (Rel. 34, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 01-NOV-1997 (Rel. 35, Last annotation update)

DE D3 PROTEIN (FRAGMENT).

GN D3.

OS Onchocerca volvulus.

OC Eukaryota; Metazoa; Nematoda; Secernentea; Spirurida;

OC Filarioidea; Onchocercidae; Onchocerca.

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE; 97045813.

RA ERTMANN K.D., GALLIN M.Y.;

RT "Onchocerca volvulus: identification of cDNAs encoding a putative

phosphatidyl-ethanolamine-binding protein and a putative partially

processed mRNA precursor.";

RL Gene 174:203-207(1996).

CC -!- SIMILARITY: BELONGS TO THE PHOSPHATIDYLETHANOLAMINE-BINDING

CC PROTEIN FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; X87989; CAA61242.1; -

DR PROSITE; PS01220; PBP; 1.

DR PFAM; PF01161; PBP; 1.

FT NON\_TER 1

SQ SEQUENCE 134 AA; 14880 MW; FBB84137 CRC32;

Query Match 100.0%; Score 15; DB 1; Length 134;

Best Local Similarity 25.0%; Pred. No. 4.56e+03;

Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 25 PGALYTLV 32

QY 2 PXXXXXXV 9

RESULT 6

ID CRYA\_EULFU STANDARD; PRT; 173 AA.

AC P02494;

DT 21-JUL-1986 (Rel. 01, Created)

DT 21-JUL-1986 (Rel. 01, Last sequence update)

DT 15-JUL-1999 (Rel. 38, Last annotation update)

DE ALPHA CRYSTALLIN A CHAIN.

GN CRYAA.

OS Eulemur fulvus fulvus (Brown lemur).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

OC Eutheria; Primates; Strepsirhini; Lemnidae; Eulemur.

RN [1]

RP PARTIAL SEQUENCE.

RA DE JONG W.W., ZWEERS A., GOODMAN M.;

RT "Trends in the molecular evolution of alpha-crystallin.";

RL (In) Peeters H. (eds.);

RL Protides of the biological fluids, Proc. 28th colloquium, pp.161-164,

RL Pergamon Press, Oxford (1980).

CC -!- FUNCTION: MAY CONTRIBUTE TO THE TRANSPARENCY AND REFRACTIVE INDEX

CC OF THE LENS.

CC -!- SIMILARITY: BELONGS TO THE SMALL HEAT SHOCK PROTEIN (HSP20)

CC FAMILY. STRONG TO ALPHA(B)-CRYSTALLIN.

DR PIR; A02897; CYLEAA.

DR PROSITE; PS01031; HSP20; 1.

DR PFAM; PF00011; HSP20; 1.

DR PFAM; PF00525; crystallin; 1.

KW Eye lens protein; Acetylation; Glycoprotein.

FT MOD\_RES 1 ACETYLATION (PROBABLE).

FT CARBOHYD 162 GLCNAC (BY SIMILARITY).

SQ SEQUENCE 173 AA; 19819 MW; BDA6E55B CRC32;

Query Match 100.0%; Score 15; DB 1; Length 173;

Best Local Similarity 25.0%; Pred. No. 4.56e+03;

Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 82 PEDLTQV 89

QY 2 PXXXXXXV 9

RESULT 7

ID DNAT\_ECOLI STANDARD; PRT; 179 AA.

AC P07904;

DT 01-AUG-1988 (Rel. 08, Created)

DT 01-AUG-1988 (Rel. 08, Last sequence update)

DT 01-NOV-1997 (Rel. 35, Last annotation update)

DE PRIMOSOMAL PROTEIN I.

GN DNAT.

OS Escherichia coli.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

OC Escherichia.

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE; 86149284.

RA MASAI H., BOND M.W., ARAI K.-I.;

RT "Cloning of the Escherichia coli gene for primosomal protein i: the

relationship to dnaat, essential for chromosomal DNA replication.";

RL Proc. Natl. Acad. Sci. U.S.A. 83:1256-1260(1986).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN-K12;

RX MEDLINE; 89008392.

RA MASAI H., ARAI K.-I.;

RT "Operon structure of dnaat and dnaC genes essential for normal and

stable DNA replication of Escherichia coli chromosome.";

RL J. Biol. Chem. 263:15083-15093(1988).

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN-K12 / MG1655;

RX MEDLINE; 95334362.

RA BURLAND V.D., PLUNKETT G. III, SOFIA H.J., DANIELS D.L.,

RA BLATTNER F.R.;

RT "Analysis of the Escherichia coli genome VI: DNA sequence of the

region from 92.8 through 100 minutes.";

RL Nucleic Acids Res. 23:2105-2119(1995).

RN [4]

RP SEQUENCE OF 108-179 FROM N.A.

RX MEDLINE; 87280100.

RA NAKAYAMA N., BOND M.W., MIYAJIMA A., KOBORI J., ARAI K.-I.;

RT "Structure of Escherichia coli dnaC. Identification of a cysteine

residue possibly involved in association with dnaB protein.";

RL J. Biol. Chem. 262:10475-10480(1987).

CC -!- FUNCTION: THIS PROTEIN IS REQUIRED FOR PRIMOSOME-DEPENDENT NORMAL

DNA REPLICATION. IT IS ALSO INVOLVED IN INDUCING STABLE DNA

REPLICATION DURING SOS RESPONSE. IT FORMS, IN CONCERT WITH DNAB

PROTEIN & OTHER PREPRIMING PROTEINS DNAC, N', N', A PREPRIMING

PROTEIN COMPLEX ON THE SPECIFIC SITE OF THE TEMPLATE DNA

CC RECOGNIZED BY PROTEIN N'.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```
RESULT 2
ID CCKN_HUMAN STANDARD; PRT; 115 AA.
AC P06307;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE PROCHOLECYSTOKININ PRECURSOR (CCK).
GN CCK.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 85166246.
RA TAKAHASHI Y., KATO K., HAYASHIZAKI Y., WAKABAYASHI T., OHTSUKA E.,
RA MATSUKI S., IKHARA M., MATSUBARA K.;
RT "Molecular cloning of the human cholecystokinin gene by use of a
RT synthetic probe containing deoxynosine."
RL Proc. Natl. Acad. Sci. U.S.A. 82:1931-1935(1985).
RN [2]
RP SEQUENCE FROM N.A.
RA KATO K., TAKAHASHI Y., MATSUBARA K.;
RT "Molecular cloning of the human cholecystokinin gene."
RL Ann. N.Y. Acad. Sci. 448:613-615(1985).
CC -1- FUNCTION: THIS PEPTIDE HORMONE INDUCES GALL BLADDER CONTRACTION
CC AND THE RELEASE OF PANCREATIC ENZYMES IN THE GUT. ITS FUNCTION
CC IN THE BRAIN IS NOT CLEAR.
CC -1- PTM: THE PRECURSOR CLEAVED BY ENZYMES TO PRODUCE A NUMBER OF
CC ACTIVE CHOLECYSTOKININS.
CC -1- SIMILARITY: BELONGS TO THE GASTRIN/CHOLECYSTOKININ FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L00354; AAA53094.1; -.
DR PIR; A01623; GMHUCP.
DR MIM; L18440; -.
DR PROSITE; PS00259; GASTRIN; 1.
DR PFAM; PF00918; Gastrin; 1.
DR Hormone; Cleavage on pair of basic residues; Signal; Amidation;
KW Sulfatation.
FT SIGNAL 1 20
FT CHAIN 21 115 PROCHOLECYSTOKININ.
FT PEPTIDE 46 103 CHOLECYSTOKININ CCK58.
FT PEPTIDE 65 103 CHOLECYSTOKININ CCK39.
FT PEPTIDE 71 103 CHOLECYSTOKININ CCK33.
FT PEPTIDE 92 103 CHOLECYSTOKININ CCK12.
FT PEPTIDE 96 103 CHOLECYSTOKININ CCK8.
FT MOD_RES 97 97 SULFATATION.
FT MOD_RES 103 103 AMIDATION (G-104 PROVIDE AMIDE GROUP).
SQ SEQUENCE 115 AA; 12669 MW; 274BD45D CRC32;

Query Match 100.0%; Score 15; DB 1; Length 115;
Best Local Similarity 25.0%; Pred. No. 4.56e+03;
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 73 PSGRMSIV 80
QY 2 PXXXXXXV 9

RESULT 3
ID ANFC_SCYCA STANDARD; PRT; 115 AA.
AC P23259;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE C-TYPE NATRIURETIC PEPTIDE (CNP-115).
OS Scyliorhinus canicula (Spotted dogfish) (Spotted catshark).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
OC Elasmobranchii; Carcharhiniformes; Scyliorhinidae; Scyliorhinus.
RN [1]
RP SEQUENCE.
RC TISSUE=HEART ATRIUM, AND HEART VENTRICLE;
RX MEDLINE; 91243822.
RA SUZUKI R., TAKAHASHI A., HAZON N., TAKEI Y.;
RT "Isolation of high-molecular-weight C-type natriuretic peptide from
RT the heart of a cartilaginous fish (European dogfish, Scyliorhinus
RT canicula).";
RL FEBS Lett. 282:321-325(1991).
CC -1- FUNCTION: VASORELAXANT ACTIVITY. HAS A CGMP-STIMULATING ACTIVITY
CC (BY SIMILARITY).
CC -1- TISSUE SPECIFICITY: CNP-115 IS DIFFERENTIALLY PROCESSED TO
CC PRODUCE CNP-38 AND CNP-39 IN THE HEART AND CNP-22 IN THE BRAIN.
CC -1- SIMILARITY: BELONGS TO THE NATRIURETIC PEPTIDES FAMILY.
DR PIR; S15822; S15822.
DR PROSITE; PS00363; NATRIURETIC_PEPTIDE; 1.
DR PFAM; PF00212; ANP; 1.
KW Vasoactive.
FT PEPTIDE 77 115 CNP-39.
FT PEPTIDE 78 115 CNP-38.
FT PEPTIDE 94 115 CNP-22.
FT DISULFID 99 115 BY SIMILARITY.
SQ SEQUENCE 115 AA; 12885 MW; 37330059 CRC32;

Query Match 100.0%; Score 15; DB 1; Length 115;
Best Local Similarity 25.0%; Pred. No. 4.56e+03;
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 95 PSRGCFGV 102
QY 2 PXXXXXXV 9

RESULT 4
ID CC42_ANOGA STANDARD; PRT; 134 AA.
AC Q17031; Q93110;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE CDC42 HOMOLOG (25 KD GTP-BINDING PROTEIN) (FRAGMENT).
GN CDC42.
OS Anopheles gambiae (African malaria mosquito).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;
OC Culicidae; Anopheles.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=G3; TISSUE=MIDGUT;
RA DIMOPOULOS G., RICHMAN A., DELLA TORRE A., RUBIO J., KAFATOS F.C.,
RA LOUIS C.;
RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE RHO FAMILY IN THE RAS SUPERFAMILY.
CC CDC42 HOMOLOG.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z69980; CAA93820.1; -.
DR HSP; P21181; IAM4.
DR PFAM; PF00071; ras; 1.
KW GTP-binding; Lipoprotein; Prenylation.
FT NON_TER 1 1
FT NP_BIND <1 4 GTP (BY SIMILARITY).
FT NP_BIND 58 61 GTP (BY SIMILARITY).
FT LIPID 131 131 GERANYL-GERANYL (BY SIMILARITY).
```

\*\*\*\*\*  
W P E L L  
\*\*\*\*\*  
(TM)  
\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd  
MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Sat Apr 15 01:56:39 2000; MasPar time 3.04 Seconds  
Tabular output not generated. 88.473 Million cell updates/sec.

Title: >US-08-452-843-27  
Description: (1-9) from US08452843.pep  
Perfect Score: 15  
Sequence: 1 XPXXXXXXV 9  
Scoring table: PAM 150  
Gap 15  
Searched: 82229 seqs, 29864866 residues  
Post-processing: Minimum Match 0%  
Listing first 45 summaries  
Database: swiss-prot38  
1:swissprot  
Statistics: Mean 11.169; Variance 8.248; scale 1.354

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

sult No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	15	100.0	46	1	CSPA_AERHY	MAJOR COLD-SHOCK PROTE	4.56e+03
2	15	100.0	115	1	CKKN_HUMAN	PROCHOLCYSTOKININ PRE	4.56e+03
3	15	100.0	115	1	ANFC_SCYCA	C-TYPE NATRIURETIC PEP	4.56e+03
4	15	100.0	134	1	CC42_ANOGA	CD42 HOMOLOG (25 KD G	4.56e+03
5	15	100.0	134	1	D3_ONCVO	D3 PROTEIN (FRAGMENT)	4.56e+03
6	15	100.0	173	1	CRAA_EULFU	ALPHA CRYSTALLIN A CHA	4.56e+03
7	15	100.0	179	1	DNAT_ECOLI	PRIMOSOMAL PROTEIN I.	4.56e+03
8	15	100.0	204	1	AR11_XENLA	APOPTOSIS REGULATOR R1	4.56e+03
9	15	100.0	210	1	CU30_BOMMO	LARVAL CUTICLE PROTEIN	4.56e+03
10	15	100.0	241	1	CRTA_RHOCA	SPHEROIDE NE MONOOXYGEN	4.56e+03
11	15	100.0	253	1	ADH_DRODI	ALCOHOL DEHYDROGENASE	4.56e+03
12	15	100.0	253	1	ADH_DROFL	ALCOHOL DEHYDROGENASE	4.56e+03
13	15	100.0	289	1	DAPA_METJA	DIHYDRODIPICOLINATE SY	4.56e+03
14	15	100.0	331	1	APAZ_KLUJA	5',5''-p-1,p-4-TETRAP	4.56e+03
15	15	100.0	339	1	CYSM_ALCEU	CYSTINE SYNTHASE (EC	4.56e+03
16	15	100.0	346	1	D3HI_RAT	3-HYDROXYISOBUTYRATE D	4.56e+03
17	15	100.0	368	1	CYHY_GLOSO	CYANIDE HYDRATASE (EC	4.56e+03
18	15	100.0	380	1	DAPA_MAIZE	DIHYDRODIPICOLINATE SY	4.56e+03
19	15	100.0	383	1	CYSL_SPIOL	CYSTINE SYNTHASE. CHL	4.56e+03
20	15	100.0	386	1	SHIB_RAT	5-HYDROXYTRYPTAMINE 1B	4.56e+03
21	15	100.0	390	1	SH1B_RABIT	5-HYDROXYTRYPTAMINE 1B	4.56e+03
22	15	100.0	441	1	AP50_CAEEL	CLATHRIN COAT ASSEMBLY	4.56e+03
23	15	100.0	455	1	A2AC_CAVPO	ALPHA-2C ADRENERGIC RE	4.56e+03

24	15	100.0	469	1	A2AC_DIDMA	ALPHA-2C ADRENERGIC RE	4.56e+03
25	15	100.0	473	1	DLDH_SYNY3	DIHYDROLIPOAMIDE DEHYD	4.56e+03
26	15	100.0	477	1	DLDH_TRYCR	DIHYDROLIPOAMIDE DEHYD	4.56e+03
27	15	100.0	481	1	ATPB_PVILI	ATP SYNTHASE BETA CHAI	4.56e+03
28	15	100.0	492	1	CAT1_LYCES	CATALASE ISOZYME 1 (EC	4.56e+03
29	15	100.0	492	1	CAT1_HORVU	CATALASE ISOZYME 1 (EC	4.56e+03
30	15	100.0	501	1	ABC1_YEAST	ABC1 PROTEIN PRECURSOR	4.56e+03
31	15	100.0	507	1	ATPA_MAIZE	ATP SYNTHASE ALPHA CHA	4.56e+03
32	15	100.0	529	1	DNB2_ADE02	EARLY E2A DNA-BINDING	4.56e+03
33	15	100.0	566	1	BGLC_MAIZE	BETA-GLUCOSIDASE. CHLO	4.56e+03
34	15	100.0	590	1	CO8B_RABIT	COMPLEMENT COMPONENT C	4.56e+03
35	15	100.0	598	1	CYLI_HUMAN	CYCLOCIN I (MULTIPLE-BA	4.56e+03
36	15	100.0	724	1	ATIL_VACCV	94 KD A-TYPE INCLUSION	4.56e+03
37	15	100.0	725	1	AREA_PENCH	NIROGEN REGULATORY PR	4.56e+03
38	15	100.0	781	1	CTNB_XENLA	BETA-CATENIN.	4.56e+03
39	15	100.0	818	1	CTNB_URECA	BETA-CATENIN.	4.56e+03
40	15	100.0	1001	1	ATCA_RABIT	CALCIUM-TRANSPORTING A	4.56e+03
41	15	100.0	1036	1	ATHL_RAT	POTASSIUM-TRANSPORTING	4.56e+03
42	15	100.0	1048	1	ANGR_VIBAN	ANGR PROTEIN.	4.56e+03
43	15	100.0	1108	1	CN3B_RAT	CGMP-INHIBITED 3',5'-C	4.56e+03
44	15	100.0	1472	1	ATC9_YEAST	PROBABLE CALCIUM-TRANS	4.56e+03
45	15	100.0	2167	1	BEM2_YEAST	GTPASE ACTIVATING PROT	4.56e+03

ALIGNMENTS

RESULT 1  
ID CSPA\_AERHY STANDARD; PRT; 46 AA.  
AC Q44078; 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 15-JUL-1998 (Rel. 36, Last annotation update)  
DE MAJOR COLD-SHOCK PROTEIN (FRAGMENT).  
GN CSPA.  
OS Aeromonas hydrophila.  
OC Bacteria; Proteobacteria; gamma subdivision; Aeromonas group;  
OC Aeromonas.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA FRANCIS K.P., STEWART G.S.A.B.;  
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.  
CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).  
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).  
CC -1- INDUCTION: IN RESPONSE TO LOW TEMPERATURE (BY SIMILARITY).  
CC -1- SIMILARITY: BELONGS TO THE COLD-SHOCK DOMAIN (CSD) FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; U60026; AAC80230.1; -  
CC HSSP; P15277; 1MJC.  
CC PROSITE; PS00352; COLD\_SHOCK; 1.  
CC PFAM; PF00313; CSD; 1.  
CC Transcription regulation; DNA-binding; Activator.  
CC NON\_TER 1 1  
CC DOMAIN <1 >46 CSD.  
CC NON\_TER 46 46  
CC SEQUENCE 46 AA; 5105 MW; C07E96EC CRC32;

Query Match 100.08; Score 15; DB 1; Length 46;  
Best Local Similarity 25.08; Pred. No. 4.56e+03;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
Db 9 PTGSKDV 16  
Qy 2 PXXXXXXV 9

THIS PAGE BLANK (USPTO)

```
##residues      88-103,'M',105-143,'E',145-184,'A',186,'E',188-266
                #label MAR
##cross-references GB:M16426; NID:g165111; PID:g165112
##note          this sequence has the d11 allotypic marker, 104-Met, and
                the e15 allotypic marker, 185-Ala
REFERENCE
#authors        A90245
#journal        Fruchter, R.G.; Jackson, S.A.; Mole, L.E.; Porter, R.R.
#journal        Biochem. J. (1970) 116:249-259
#title          Sequence studies of the Fd section of the heavy chain of
                rabbit immunoglobulin G.
##cross-references MURID:70110015
#accession      A90245
##molecule_type protein
##residues      132-143,'E',145-161 #label FRU
REFERENCE
#authors        A94416
#book           Hill, R.L.; Lebovitz, H.E.; Fellows Jr., R.E.; Delaney, R.
                in Gamma Globulins, Nobel Symp. 3, Killander, J., ed.,
                pp.109-127, Almquist and Wiksell, Stockholm, 1967
#accession      A94416
##molecule_type protein
##residues      129-131,155-172,'D',174-184,'A',186,'E',188-200,'D',
                202-217,'E',219-232,'Q',234-245,'D',247-255,'G',
                257-259,'D',261-265,'D',267-279,'W',281-283,'S',
                285-322 #label HIL
##note          this has the e15 allotypic marker, 185-Ala
                An immunoglobulin heterotetramer subunit consists of two
                identical light (kappa or lambda) and two identical heavy
                (alpha, delta, epsilon, gamma, or mu) chains usually
                stabilized by interchain disulfide bonds. In some cases,
                such as IgA and IgM, the subunits associate into larger
                oligomers.
CLASSIFICATION  #superfamily immunoglobulin C region; immunoglobulin homology
KEYWORDS        duplication; glycoprotein; heterotetramer; immunoglobulin
FEATURE
20-82           #domain immunoglobulin homology #label IGG1\
130-199         #domain immunoglobulin homology #label IGG2\
236-303         #domain immunoglobulin homology #label IGG3\
173            #binding_site carbohydrate (Asn) (covalent) #status
                predicted
SUMMARY          #length 323 #molecular-weight 35404 #checksum 1467

Query Match      100.0%; Score 15; DB 1; Length 323;
Best Local Similarity 25.0%; Pred. No. 4.88e+03;
Matches          2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
```

```
Db 133 PEVTCVVV 140
   |
QY 2 PXXXXXXV 9
```

Search completed: Sat Apr 15 01:56:21 2000  
Job time : 18 secs.



```

27-57,73-217      #product sonatotropin 2, short form #status predicted
                  #label SOS\
79-191,208-215    #disulfide_bonds #status predicted\
166               #binding_site carbohydrate (Asn) (covalent) #status
                  predicted
SUMMARY           #length 217 #molecular-weight 24999 #checksum 6227

Query Match      100.0%; Score 15; DB 1; Length 217;
Best Local Similarity 25.0%; Pred. No. 4.88e+03;
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 115 PVQLRSV 122
|
QY 2 PXXXXXXV 9

RESULT 13
ENTRY SOMS #type complete
TITLE parotid secretory protein precursor - mouse
ALTERNATE_NAMES PSP
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE 31-Dec-1998 #sequence_revision 31-Dec-1998 #text_change
20-Mar-1998
ACCESSIONS A23031; 153236
REFERENCE A23031
#authors Madsen, H.O.; Hjorth, J.P.
#journal Nucleic Acids Res. (1985) 13:1-13
#title Molecular cloning of mouse PSP mRNA.
#cross-references MUID:85215456
#accession A23031
#molecule_type mRNA
#residues 1-235 #label MAD
#cross-references GB:X01697; NID:g53810; PID:g758163
REFERENCE I53236
#authors Poulsen, K.; Jakobsen, B.K.; Mikkelsen, B.M.; Harmark, K.;
Nielsen, J.T.; Hjorth, J.P.
#journal EMBO J. (1986) 5:1891-1896
#title Coordination of murine parotid secretory protein and salivary
amylase expression.
#cross-references MUID:87004556
#accession I53236
#status preliminary; translated from GB/EMBL/DBJ
#molecule_type DNA
#residues 1-87 #label RES
#cross-references GB:M26807; NID:g200556; PID:g554264
COMMENT PSP is the most abundant protein in the parotid gland. Its function
is not known; however, its production is coordinated with that of
salivary amylase.

GENETICS
#gene Psp
#map_position 2
#introns 41/1
#note list of introns may be incomplete
CLASSIFICATION #superfamily parotid secretory protein.
KEYWORDS parotid gland; saliva
FEATURE
1-20 #domain signal sequence #status predicted #label SIG\
21-235 #product parotid secretory protein #status predicted
#label MAT
SUMMARY #length 235 #molecular-weight 24753 #checksum 1500

Query Match 100.0%; Score 15; DB 1; Length 235;
Best Local Similarity 25.0%; Pred. No. 4.88e+03;
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 43 PQNLNDV 50
|
QY 2 PXXXXXXV 9

RESULT 14
ENTRY DNPDPW #type complete
TITLE repeat element protein - Campoletis sonorensis virus

```

```

ORGANISM #formal_name Campoletis sonorensis virus, Csv
#note host Campoletis sonorensis (parasitic wasp); Heliiothis
virescens
DATE 31-Mar-1990 #sequence_revision 31-Mar-1990 #text_change
21-Nov-1997
ACCESSIONS A31823
REFERENCE A31823
#authors Theilmann, D.A.; Summers, M.D.
#journal Virology (1988) 167:329-341
#title Identification and comparison of Campoletis sonorensis virus
transcripts expressed from four genomic segments in the
insect hosts Campoletis sonorensis and Heliiothis virescens.
#cross-references MUID:89073734
#accession A31823
#molecule_type mRNA
#residues 1-235 #label THE
#cross-references GB:M23437; GB:M16998; NID:g232408; PID:g323409
COMMENT The genome of this virus consists of at least 28 closed circular
superhelical DNA segments; three of them contain homologous DNA
sequences that code for one or several tandem-repeated element
proteins.
CLASSIFICATION #superfamily parasitic wasp virus repeat element protein
FEATURE
57-235 #domain repeat element #label RPE
SUMMARY #length 235 #molecular-weight 28044 #checksum 4181

Query Match 100.0%; Score 15; DB 1; Length 235;
Best Local Similarity 25.0%; Pred. No. 4.88e+03;
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 25 PLKMISV 32
|
QY 2 PXXXXXXV 9

RESULT 15
ENTRY GRRB #type complete
TITLE Ig gamma chain C region - rabbit
ORGANISM #formal_name Oryctolagus cuniculus #common_name domestic
rabbit
DATE 24-Apr-1984 #sequence_revision 15-Nov-1984 #text_change
20-Mar-1998
ACCESSIONS A91749; A90290; A93928; A90245; A94416; A02161
REFERENCE A91749
#authors Bernstein, K.E.; Alexander, C.B.; Mage, R.G.
#journal Immunogenetics (1983) 18:387-397
#title Nucleotide sequence of a rabbit IgG heavy chain from the
recombinant P-I haplotype.
#cross-references MUID:84030930
#accession A91749
#molecule_type mRNA
#residues 1-323 #label BER
#note this sequence has the d12 allotypic marker, 104-Thr, and
the e14 marker, 185-Thr
REFERENCE A90290
#authors Pratt, D.M.; Mole, L.E.
#journal Biochem. J. (1975) 151:337-349
#title Sequence studies on the constant region of the Fd sections of
rabbit immunoglobulin G of different allotype.
#cross-references MUID:76135469
#accession A90290
#molecule_type protein
#residues 1-47,'E',49-71,'PV',72-128 #label PRA
REFERENCE A93928
#authors Martens, C.L.; Moore, K.W.; Steinmetz, M.; Hood, L.; Knight,
K.L.
#journal Proc. Natl. Acad. Sci. U.S.A. (1982) 79:6018-6022
#title Heavy chain genes of rabbit IgG: isolation of a cDNA encoding
gamma heavy chain and identification of two genomic C-gamma
genes.
#cross-references MUID:83299917
#accession A93928
#molecule_type mRNA

```

```

Query Match      100.0%; Score 15; DB 1; Length 211;
Best Local Similarity 25.0%; Pred. NO. 4.88e+03;
Matches          2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 70 PLPNRMNV 77
Qy 2 PXXXXXXV 9

RESULT 11
ENTRY RBYD #type complete
TITLE dihydrofolate reductase (EC 1.5.1.3) - yeast (Saccharomyces cerevisiae)
ALTERNATE_NAMES protein O5231; protein YOR236w
ORGANISM #formal_name Saccharomyces cerevisiae
DATE 30-Sep-1990 #sequence_revision 30-Sep-1990 #text_change 12-Dec-1997
ACCESSIONS JT0269; JT0274; S06312; S67129
REFERENCE A91592
#authors Fling, M.E.; Kopf, J.; Richards, C.A.
#journal Gene (1988) 63:165-174
#title Nucleotide sequence of the dihydrofolate reductase gene of Saccharomyces cerevisiae.
#cross-references EMBL:88255864
#accession JT0269
#molecule_type DNA
#residues 1-211 #label FLI
#cross-references GB:M18578; EMBL:M26667; NID:gl71396; PID:gl71397
REFERENCE A91593
#authors Barclay, B.J.; Huang, T.; Nagel, M.G.; Misener, V.L.; Game, J.C.; Wahl, G.M.
#journal Gene (1988) 63:175-185
#title Mapping and sequencing of the dihydrofolate reductase gene (DFR1) of Saccharomyces cerevisiae.
#cross-references MUID:88255865
#accession JT0274
#molecule_type DNA
#residues 1-211 #label BAR
#cross-references EMBL:M26668; NID:g295603; PID:g295604
REFERENCE S06312
#authors Lagosky, P.A.; Taylor, G.R.; Haynes, R.H.
#journal Nucleic Acids Res. (1987) 15:10355-10371
#title Molecular characterization of the Saccharomyces cerevisiae dihydrofolate reductase gene (DFR1).
#cross-references MUID:88096572
#accession S06312
#molecule_type DNA
#residues 1-211 #label LAG
#cross-references EMBL:Y00887
#note the authors translated the codon GTA for residue 27 as Leu; the sequence shown follows the authors' translation
REFERENCE S67104
#authors Boyer, J.; Fairhead, C.; Gaillon, L.; Galisson, F.; Michaux, G.; Thierry, A.; Dujon, B.
#submission submitted to the Protein Sequence Database, July 1996
#accession S67129
#molecule_type DNA
#residues 1-211 #label BOY
#cross-references EMBL:Z75144; NID:g1420540; PID:e252096; PID:g1420541; MIPS:YOR236w
#experimental_source strain S288C
COMMENT This enzyme catalyzes the NADPH-dependent reduction of dihydrofolate to tetrahydrofolate.
GENETICS
#gene SGD:DFR1
#map_position 15R
#cross-references SGD:S0005762; MIPS:YOR236w
CLASSIFICATION #superfamily type I dihydrofolate reductase; type I dihydrofolate reductase homology
KEYWORDS NADP; oxidoreductase
FEATURE 8-132 #domain type I dihydrofolate reductase homology #label DFR1
34,38,68,74 #binding_site substrate (Glu, Phe, Arg) #status predicted
#length 211 #molecular-weight 24261 #checksum 8536
SUMMARY

```

---

```

Query Match      100.0%; Score 15; DB 1; Length 211;
Best Local Similarity 25.0%; Pred. NO. 4.88e+03;
Matches          2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 70 PLPNRMNV 77
Qy 2 PXXXXXXV 9

RESULT 12
ENTRY STHUV #type complete
TITLE somatotropin 2 precursor - human
ALTERNATE_NAMES growth hormone 2; growth hormone variant; hGH-V; placental somatotropin
CONTAINS somatotropin 2, long form; somatotropin 2, short form
ORGANISM #formal_name Homo sapiens #common_name man
DATE 17-Dec-1982 #sequence_revision 10-Feb-1995 #text_change 05-Jun-1998
ACCESSIONS D32435; B28072; A01511; I52104
REFERENCE A32435
#authors Chen, E.Y.; Liao, Y.C.; Smith, D.H.; Barrera-Saldana, H.A.; Gellinas, R.E.; Seeburg, P.H.
#journal Genomics (1989) 4:479-497
#title The human growth hormone locus: nucleotide sequence, biology, and evolution.
#cross-references MUID:89307277
#accession D32435
#molecule_type DNA
#residues 1-217 #label CHE
#cross-references GB:J03071; NID:gl83148; PID:gl83152
REFERENCE A92725
#authors Cooke, N.E.; Ray, J.; Emery, J.G.; Liebhauer, S.A.
#journal J. Biol. Chem. (1988) 263:9001-9006
#title Two distinct species of human growth hormone-variant mRNA in the human placenta predict the expression of novel growth hormone proteins.
#cross-references MUID:88243769
#accession B28072
#molecule_type mRNA
#residues 1-217 #label COO
REFERENCE A01511
#authors Seeburg, P.H.
#journal DNA (1982) 1:239-249
#title The human growth hormone gene family: nucleotide sequences show recent divergence and predict a new polypeptide hormone
#cross-references MUID:83182010
#accession A01511
#molecule_type DNA
#residues 1-34,'P',36-217 #label SEE
REFERENCE I52104
#authors Ignotz, A.; Scippo, M.L.; Frankenne, F.; Hennen, G.
#journal Arch. Int. Physiol. Biochim. (1988) 96:63-67
#title Cloning and nucleotide sequence of placental hGH-V cDNA.
#cross-references MUID:8904984
#accession I52104
#status preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues 1-217 #label RES
#cross-references GB:M38451; NID:gl83179; PID:gl83180
COMMENT The gene for this hormone appears to be transcribed in the placenta.
GENETICS
#gene GDB:GH2
#cross-references GDB:119983; OMIM:139240
#map_position 17q22-17q24
#introns 4/1; 57/3; 97/3; 152/3
CLASSIFICATION #superfamily prolactin
KEYWORDS alternative splicing; glycoprotein; hormone; placenta
FEATURE 1-26 #domain signal sequence #status predicted #label SIG\
27-217 #product somatotropin 2, long form #status predicted #label SOL\

```

```

#accession S6557
##molecule_type protein
##residues 19-22;138-140,'Q';158-161 ##label FAN
CLASSIFICATION #superfamily type I dihydrofolate reductase; type I
                dihydrofolate reductase homology
KEYWORDS NADP; oxidoreductase
FEATURE
30,34,64,70 #domain type I dihydrofolate reductase homology #label
DPR\
#binding_site substrate (Glu, Phe, Asn, Arg) #status
predicted
SUMMARY #length 189 #molecular-weight 21650 #checksum 1926

Query Match 100.0%; Score 15; DB 1; Length 189;
Best Local Similarity 25.0%; Pred. No. 4.88e+03;
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 128 PINHRLFV 135
QY 2 PXXXXXXV 9

RESULT 10
ENTRY A29413 #type complete
TITLE ubiquinol--cytochrome-c reductase (EC 1.10.2.2) iron-sulfur
protein - Paracoccus denitrificans
ALTERNATE_NAMES complex III iron-sulfur protein; cytochrome bc1 complex
iron-sulfur protein; Rieske iron-sulfur protein
ORGANISM #formal_name Paracoccus denitrificans
DATE 31-Mar-1989 #sequence_revision 31-Mar-1989 #text_change
17-Mar-1999
ACCESSIONS A29413; S65367
REFERENCE A92613
#authors Kuroski, B.; Ludwig, B.
#journal J. Biol. Chem. (1987) 262:13805-13811
#title The genes of the Paracoccus denitrificans bc-1 complex.
Nucleotide sequence and homologies between bacterial and
mitochondrial subunits.
#cross-references MUID:88007612
#accession A29413
##molecule_type DNA
##residues 1-190 #label KUR
##cross-references GB:M1752; NID:g150569; PID:g294081
REFERENCE S65367
#authors Kleymann, G.; Iwata, S.; Wiesmueller, K.H.; Ludwig, B.;
Haase, W.; Michel, H.
#journal Eur. J. Biochem. (1995) 230:359-363
#title Immunoelectron microscopy and epitope mapping with monoclonal
antibodies suggest the existence of an additional
N-terminal transmembrane helix in the cytochrome b subunit
of bacterial ubiquinol:cytochrome-c oxidoreductases.
#cross-references MUID:95324547
#accession S65367
##status preliminary
##molecule_type protein
##residues 17-23;37-50;51,'X';52-60 #label KLE
CLASSIFICATION #superfamily ubiquinol--cytochrome-c reductase iron-sulfur
protein
KEYWORDS 2Fe-2S; electron transfer; membrane-associated complex;
metalloprotein; oxidoreductase; respiratory chain; Rieske
iron-sulfur protein
FEATURE
132,134,152,155 #binding_site 2Fe-2S cluster (Cys, His, Cys, His)
(covalent) #status predicted\
137-154 #disulfide_bonds #status predicted\
155 #active_site His #status predicted
SUMMARY #length 190 #molecular-weight 20299 #checksum 7756

Query Match 100.0%; Score 15; DB 2; Length 190;
Best Local Similarity 25.0%; Pred. No. 4.88e+03;
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 171 PQNLHPIV 178

```

ORGANISM #formal\_name Rattus norvegicus #common\_name Norway rat  
DATE 10-Aug-1990 #sequence\_revision 10-Aug-1990 #text\_change 26-Feb-1999

ACCESSIONS A35209; S12724; S04070; S04593; S14190; S65373

REFERENCE Yamada, M.; Amuro, N.; Goto, Y.; Okazaki, T.  
J. Biol. Chem. (1990) 265:7687-7692  
#title Structural organization of the rat cytochrome c oxidase subunit IV gene.  
#cross-references MUID:90237079

#accession A35209 preliminary  
##status  
##residues 1-169 #label YAM  
##cross-references GB:J05425; NID:g203516; PID:g203517

REFERENCE S12724  
Amuro, N.; Yamada, M.; Goto, Y.; Okazaki, T.  
Nucleic Acids Res. (1990) 18:3992  
#title Complete nucleotide sequence of the gene encoding rat cytochrome c oxidase subunit IV.  
#cross-references MUID:90326528

#accession S12724 preliminary  
##status  
##molecule\_type DNA  
##residues 1-169 #label AMU  
##cross-references EMBL:J05425; NID:g203516; PID:g203517

REFERENCE S04070  
Goto, Y.; Amuro, N.; Okazaki, T.  
Nucleic Acids Res. (1989) 17:2851  
#title Nucleotide sequence of cDNA for rat brain and liver cytochrome c oxidase subunit IV.  
#cross-references MUID:89240039

#accession S04070  
##molecule\_type mRNA  
##residues 1-169 #label GOT  
##cross-references EMBL:X14209; NID:g55989; PID:g55990

REFERENCE S04593  
Gopalan, G.; Droste, M.; Kadenbach, B.  
Nucleic Acids Res. (1989) 17:4376  
#title Nucleotide sequence of cDNA encoding subunit IV of cytochrome c oxidase from fetal rat liver.  
#cross-references MUID:89296488

#accession S04593  
##molecule\_type mRNA  
##residues 1-169 #label GOP  
##cross-references EMBL:X15029; NID:g55980; PID:g55981

REFERENCE S14190  
Virbasius, J. V.; Scarpulla, R. C.  
Nucleic Acids Res. (1990) 18:6581-6586  
#title The rat cytochrome c oxidase subunit IV gene family: tissue-specific and hormonal differences in subunit IV and cytochrome c mRNA expression.  
#cross-references MUID:91067442

#accession S14190  
##status  
##molecule\_type mRNA  
##residues 1-169 #label VIR  
##cross-references EMBL:X54081; NID:g57030; PID:g57031  
#note #experimental\_source strain Sprague Dawley  
#title the nucleotide sequence was submitted to the EMBL Data Library, July 1990

REFERENCE S65372  
Schaeffer, H.; Noack, H.; Halangk, W.; Brandt, U.; von Jagow, G.  
Eur. J. Biochem. (1995) 230:235-241  
#title Cytochrome-c oxidase in developing rat heart. Enzymic properties and amino-terminal sequences suggest identity of the fetal heart and the adult liver isoform.

#accession S65373 preliminary  
##status  
##molecule\_type protein  
##residues 23-45 #label SCH

GENETICS

RCO4-1  
#gene  
#introns 25/1; 81/1; 125/1  
FUNCTION  
#description the cytochrome-c oxidase complex catalyzes the oxidation of four molecules of reduced cytochrome c in the intracristal (or intermembrane) space using one oxygen molecule and four protons from the mitochondrial matrix producing two molecules of water and lowering the concentration of protons in the mitochondrial matrix  
#pathway oxidative phosphorylation; respiratory chain  
CLASSIFICATION #superfamily cytochrome-c oxidase chain IV  
KEYWORDS electron transfer; membrane-associated complex; mitochondrial inner membrane; mitochondrion; oxidative phosphorylation; oxidoreductase; respiratory chain; transmembrane protein

FEATURE 1-22  
#domain transit peptide (mitochondrion) #status predicted #label TNP  
23-169 #product cytochrome-c oxidase chain IV #status experimental #label MAR  
77-103 #domain transmembrane helix #status predicted #label TRO1

SUMMARY #length 169 #molecular-weight 19514 #checksum 1878

Query Match 100.0%; Score 15; DB 1; Length 169;  
Best Local Similarity 25.0%; Pred. No. 4.88e+03;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 45 PLPDVAHV 52  
| |  
Qy 2 PXXXXXXV 9

RESULT 7  
ENTRY #type complete  
TITLE B11 protein - tomato golden mosaic virus  
ORGANISM #formal\_name tomato golden mosaic virus  
#note host Nicotiana sp. (tobacco)  
DATE 28-Aug-1985 #sequence\_revision 28-Aug-1985 #text\_change 08-Apr-1994  
ACCESSIONS A04169  
REFERENCE A04163  
Hamilton, W.D.O.; Stein, V.E.; Coutts, R.H.A.; Buck, K.W.  
EMBO J. (1984) 3:2197-2205  
#authors Complete nucleotide sequence of the infectious cloned DNA components of tomato golden mosaic virus: potential coding regions and regulatory sequences.  
#accession A04169  
#molecule\_type DNA  
#residues 1-184 #label HAM  
COMMENT The genome consists of two circular, single-stranded DNA components, DNA A and DNA B. There are six potential coding regions, four in DNA A and two in DNA B. This protein is coded by DNA B.

GENETICS  
#map\_position segment B  
CLASSIFICATION #superfamily tomato golden mosaic virus B11 protein  
SUMMARY #length 184 #molecular-weight 21122 #checksum 6350

Query Match 100.0%; Score 15; DB 1; Length 184;  
Best Local Similarity 25.0%; Pred. No. 4.88e+03;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 65 PINATGSV 72  
|  
Qy 2 PXXXXXXV 9

RESULT 8  
ENTRY #type complete  
TITLE RDHV75  
ORGANISM dihydrofolate reductase (EC 1.5.1.3) (antifolate-resistant variant) - Chinese hamster  
#formal\_name Crictetus griseus #common\_name Chinese hamster  
DATE 31-Mar-1989 #sequence\_revision 31-Mar-1989 #text\_change ..

#authors Levy, N.S.; Malipiero, U.V.; Lebecque, S.G.; Gearhart, P.J.  
#journal J. Exp. Med. (1989) 169:2007-2019  
#title Early onset of somatic mutation in immunoglobulin VH genes.  
#cross-references MUID:89279149  
#accession JN0501  
#status translation not shown  
#molecule\_type mRNA  
#residues 1-98 #label LEV  
#experimental\_source strain BALB/cJ  
#note this sequence belongs to the VH7183 subfamily  
CLASSIFICATION #superfamily immunoglobulin V region; immunoglobulin homology  
KEYWORDS heterotetramer; immunoglobulin  
FEATURE  
15-98  
22-96  
SUMMARY #domain immunoglobulin homology #label IMM  
#disulfide\_bonds #status predicted  
#length 98 #molecular-weight 11007 #checksum 5299

Query Match 100.0%; Score 15; DB 1; Length 98;  
Best Local Similarity 25.0%; Pred. No. 4.88e+03;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 41 PEKRLWV 48  
|  
QY 2 PXXXXXXV 9

RESULT 3  
ENTRY CCPH55 #type complete  
TITLE cytochrome c555 - Prosthecochloris aestuarii  
ORGANISM #formal\_name Prosthecochloris aestuarii  
DATE 13-Jul-1981 #sequence\_revision 13-Jul-1981 #text\_change 15-Jan-1999

ACCESSIONS A00117  
REFERENCE A00116  
#authors Van Beeumen, J.; Ambler, R.P.; Meyer, T.E.; Kamen, M.D.;  
Olson, J.M.; Shaw, E.K.  
#journal Biochem. J. (1976) 159:757-774  
#title The amino acid sequences of the cytochromes c-555 from two green sulphur bacteria of the genus Chlorobium.  
#cross-references MUID:77087088  
#accession A00117  
#molecule\_type protein  
#residues 1-99 #label VAN  
#note the source is designated as Chlorobium limicola  
REFERENCE A38042  
#authors Olson, J.M.  
#journal Int. J. Syst. Bacteriol. (1978) 28:128-129  
#contents annotation; taxonomy  
CLASSIFICATION #superfamily cytochrome c6; cytochrome c6 homology  
KEYWORDS chromoprotein; electron transfer; heme; iron; photosynthesis  
FEATURE  
13-94 #domain cytochrome c6 homology #label CYC  
23-26 #binding\_site heme (Cys) (covalent) #status predicted  
27,73 #binding\_site heme iron (His, Met) (axial ligands)  
#status predicted  
SUMMARY #length 99 #molecular-weight 10473 #checksum 3605

Query Match 100.0%; Score 15; DB 1; Length 99;  
Best Local Similarity 25.0%; Pred. No. 4.88e+03;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 80 PDLTKQV 87  
|  
QY 2 PXXXXXXV 9

RESULT 4  
ENTRY HVM557 #type complete  
TITLE Ig heavy chain precursor V region (5-76) - mouse  
ORGANISM #formal\_name Mus musculus #common\_name house mouse  
DATE 30-Jun-1990 #sequence\_revision 30-Jun-1990 #text\_change 31-Mar-1997

ACCESSIONS JN0506  
REFERENCE JN0501  
#authors Levy, N.S.; Malipiero, U.V.; Lebecque, S.G.; Gearhart, P.J.  
#journal J. Exp. Med. (1989) 169:2007-2019  
#title Early onset of somatic mutation in immunoglobulin VH genes during the primary immune response.  
#cross-references MUID:89279149  
#accession JN0506  
#status translation not shown  
#molecule\_type mRNA  
#residues 1-117 #label LEV  
#experimental\_source strain BALB/cJ  
#note this sequence belongs to the VH7183 subfamily  
CLASSIFICATION #superfamily immunoglobulin V region; immunoglobulin homology  
KEYWORDS heterotetramer; immunoglobulin  
FEATURE  
1-19 #domain signal sequence #status predicted #label SIG  
20-117 #product Ig heavy chain V region (5-76) #status predicted #label IMM  
34-117 #domain immunoglobulin homology #label IMM  
41-115 #disulfide\_bonds #status predicted  
SUMMARY #length 117 #molecular-weight 12991 #checksum 8493

Query Match 100.0%; Score 15; DB 1; Length 117;  
Best Local Similarity 25.0%; Pred. No. 4.88e+03;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 60 PEKRLWV 67  
|  
QY 2 PXXXXXXV 9

RESULT 5  
ENTRY WMS14 #type complete  
TITLE submandibular gland 14K protein - mouse  
ORGANISM #formal\_name Mus musculus #common\_name house mouse  
DATE 28-Aug-1985 #sequence\_revision 28-Aug-1985 #text\_change 02-Jul-1996

ACCESSIONS A03299  
REFERENCE A93503  
#authors Windass, J.D.; Mullins, J.J.; Beecroft, L.J.; George, H.;  
Meacock, P.A.; Williams, B.R.G.; Brammar, W.J.  
#journal Nucleic Acids Res. (1984) 12:1361-1376  
#title Molecular cloning of cDNAs from androgen-independent mRNA species of DBA/2 mouse sub-maxillary glands.  
#cross-references MUID:84144035  
#accession A03299  
#molecule\_type mRNA  
#residues 1-129 #label WIN  
#note the authors translated the codon AAT for residues 104 and 124 as Asp

COMMENT Lack of a hydrophobic region at the amino end may indicate that this protein is intracellular. However, similarity to a related human sequence in entry SQHAC suggests that the amino end shown here may not be exact.

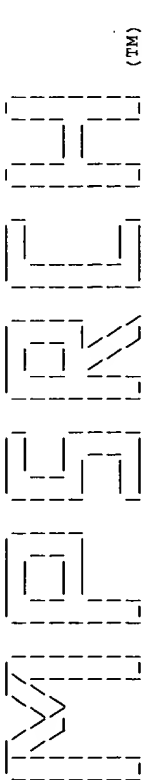
CLASSIFICATION #superfamily submandibular gland 14K protein  
KEYWORDS submandibular gland  
FEATURE  
48-74,72-106 #disulfide\_bonds #status predicted  
SUMMARY #length 129 #molecular-weight 14870 #checksum 5481

Query Match 100.0%; Score 15; DB 1; Length 129;  
Best Local Similarity 25.0%; Pred. No. 4.88e+03;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 18 PLLIEDV 25  
|  
QY 2 PXXXXXXV 9

RESULT 6  
ENTRY A35209 #type complete  
TITLE cytochrome-c oxidase (EC 1.9.3.1) chain IV precursor - rat

\*\*\*\*\*



\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Sat Apr 15 01:56:03 2000; MasPar time 3.04 Seconds  
Tabular output not generated. 118.758 Million cell updates/sec

Title: >US-08-452-843-27  
Description: (1-9) from US08452843.pep  
Perfect Score: 15  
Sequence: 1 XPXXXXXXV 9

Scoring table: PAM 150  
Gap 15

Searched: 122810 seqs, 40068593 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: pir62  
1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 10.476; Variance 8.276; scale 1.266

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	15	100.0	85	1	CCRF22 cytochrome c2 - Rhodo	4.88e+03
2	15	100.0	98	1	HVMS96 Ig heavy chain v regi	4.88e+03
3	15	100.0	99	1	CCPR55 cytochrome c55 - Pro	4.88e+03
4	15	100.0	117	1	HVMS57 Ig heavy chain precu	4.88e+03
5	15	100.0	129	1	WMMS14 submandibular gland 1	4.88e+03
6	15	100.0	169	1	A35209 cytochrome-c oxidase	4.88e+03
7	15	100.0	184	1	QOCVLG BL1 protein - tomato	4.88e+03
8	15	100.0	186	1	RDXH75 dihydrofolate reducta	4.88e+03
9	15	100.0	189	1	RDCRD dihydrofolate reducta	4.88e+03
10	15	100.0	190	2	A29413 ubiquinol--cytochrome	4.88e+03
11	15	100.0	211	1	RBDYD dihydrofolate reducta	4.88e+03
12	15	100.0	217	1	STHUV somatotropin 2 precu	4.88e+03
13	15	100.0	235	1	SQMS parotid secretory pro	4.88e+03
14	15	100.0	235	1	DNPDW repeat element protei	4.88e+03
15	15	100.0	323	1	GHRB Ig gamma chain C regi	4.88e+03
16	15	100.0	324	1	QGVZH3 H3 protein - vaccinia	4.88e+03
17	15	100.0	329	1	QVXR22 glycoprotein VP7 prec	4.88e+03
18	15	100.0	333	1	A31998 electron transfer fla	4.88e+03
19	15	100.0	334	1	WYEC tryptophan--trna liga	4.88e+03
20	15	100.0	350	1	QGB563 glycoprotein gp63 - s	4.88e+03
21	15	100.0	378	1	QBY33 ox13 intron 3 protein	4.88e+03
22	15	100.0	385	1	CBNC ubiquinol--cytochrome	4.88e+03
23	15	100.0	386	2	S52035 probable alcohol dehy	4.88e+03

24	15	100.0	387	1	ERADN1 41k fiber protein - h	4.88e+03
25	15	100.0	387	1	CBASN ubiquinol--cytochrome	4.88e+03
26	15	100.0	403	1	S53477 IMP dehydrogenase (EC	4.88e+03
27	15	100.0	412	1	RNECTA trna adenylittransfer	4.88e+03
28	15	100.0	429	1	FOLTCN gag polyprotein - hum	4.88e+03
29	15	100.0	457	1	HLMSP3 poliovirus receptor h	4.88e+03
30	15	100.0	475	1	YWBO tryptophan--trna liga	4.88e+03
31	15	100.0	478	1	S39590 paired box transcript	4.88e+03
32	15	100.0	479	1	SL5031 format-dependent nit	4.88e+03
33	15	100.0	500	1	FOVWH4 gag polyprotein - hum	4.88e+03
34	15	100.0	507	1	QVBE41 EGLF1 protein - human	4.88e+03
35	15	100.0	521	1	VGVVDH envelope glycoprotein	4.88e+03
36	15	100.0	529	1	VGNVAC major envelope glycop	4.88e+03
37	15	100.0	532	1	A40876 dimethylalliline monoo	4.88e+03
38	15	100.0	546	1	SL1180 trna adenylittransfer	4.88e+03
39	15	100.0	585	1	SDADH5 peripentonal hexon-as	4.88e+03
40	15	100.0	586	1	VGVNBF nonstructural glycopr	4.88e+03
41	15	100.0	587	1	SYHDAE 5-aminolevulinatate syn	4.88e+03
42	15	100.0	606	1	QXBO5M NADH dehydrogenase (u	4.88e+03
43	15	100.0	894	1	FAHUA2 alpha-actinin 2 - hum	4.88e+03
44	15	100.0	982	1	VCLJVS env polyprotein precu	4.88e+03
45	15	100.0	3066	1	JQ1662 genome polyprotein -	4.88e+03

ALIGNMENTS

RESULT 1  
ENTRY CCRFG2 #type complete  
TITLE cytochrome c2 - Rhodocyclus gelatinosus  
ORGANISM #formal\_name Rhodocyclus gelatinosus  
DATE 24-Apr-1984 #sequence\_revision 24-Apr-1984 #text\_change 09-Apr-1998  
ACCESSION A00089  
REFERENCE A93207  
#authors Ambler, R.P.; Meyer, T.E.; Kamen, M.D.  
#journal Nature (1979) 278:661-662  
#title Anomalies in amino acid sequences of small cytochromes c and cytochromes c' from two species of purple photosynthetic bacteria.  
#cross-references MUID:79199668  
#accession A00089  
#molecule\_type protein  
#residues 1-85 #label AMB  
COMMENT This sequence is more closely related to the sequences of cytochrome c551 from Pseudomonas and Azotobacter than to the sequences of cytochrome c2 from other species of Rhodopseudomonas.

CLASSIFICATION #superfamily cytochrome c6; cytochrome c6 homology  
KEYWORDS chromoprotein; electron transfer; heme; iron; photosynthesis  
FEATURE  
1-81 #domain cytochrome c6 homology #label CYC\  
12,15 #binding\_site heme (Cys) (covalent) #status predicted\  
16,61 #binding\_site heme iron (His, Met) (axial ligands)  
#status predicted

SUMMARY #length 85 #molecular-weight 8899 #checksum 4949

Query Match 100.0%; Score 15; DB 1; Length 85;  
Best Local Similarity 25.0%; Pred. No. 4.88e+03;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

DB 41 PALMAERV 48  
QY 2 PXXXXXXV 9

RESULT 2  
ENTRY HVMS96 #type complete  
TITLE Ig heavy chain V region (6.96) - mouse  
ORGANISM #formal\_name Mus musculus #common\_name house mouse  
DATE 30-Jun-1990 #sequence\_revision 30-Jun-1990 #text\_change 31-Mar-1997  
ACCESSION JT0501  
REFERENCE JT0501

**THIS PAGE BLANK (USPTO)**

QY 2 PXXXXXXA 9

# RESULT 13

ID R1581 standard; protein; 107 AA.  
 AC R1581;  
 DT 17-JUN-1991 (first entry)  
 DE Macrocyclic FK-506 receptor protein.  
 KW Immunosuppressant; cyclosporin A; isomerase.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT peptide 1..40  
 FT /note= "region common to human and bovine  
 FT receptor"  
 PN WO9104321-A.  
 PD 04-APR-1991.  
 PF 23-SEP-1990; U05449.  
 PR 23-SEP-1989; US-412088.  
 PR 16-JAN-1990; US-464978.  
 PA (HARD ) HARVARD COLLEGE.  
 PA (UYA-) YALE UNIV.  
 PI Schreiber SL, Harding MW;  
 DR WPI; 91-117512/16.  
 PT Receptor for FK-506 - which is inhibitor of isomerase activity of  
 PT binding proteins of both human and bovine origin  
 PS Claim 7; page 8; 14pp; English.  
 CC This receptor is specific for the immunosuppressant macrocycle FK-  
 CC 506. It does not cross react with antisera to cyclosporin A and it  
 CC exhibits isomerase activity, specifically trans peptidyl-propyl  
 CC isomerase activity. It is therefore useful in enzyme inhibitor  
 CC assays and receptor binding assays, particularly in screening for  
 CC new cpds. with immunosuppressive activity.  
 SQ Sequence 107 AA;

Query Match 100.0%; Score 12; DB 1; Length 107;  
 Best Local Similarity 25.0%; Pred. No. 7.87e+03;  
 Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
 Db 88 PGIIPPHA 95  
 QY 2 PXXXXXXA 9

# RESULT 14

ID R27966 standard; Protein; 146 AA.  
 AC R27966;  
 DT 15-MAR-1993 (first entry)  
 DE bFGF mutein BFMA.  
 KW Mutein; basic fibroblast growth factor; bFGF; disulphide bond;  
 KW security; stabilisation; site-directed mutagenesis.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT misc\_difference 69  
 FT /note= "Mutated amino acid"  
 FT misc\_difference 75  
 FT /note= "Mutated amino acid"  
 FT misc\_difference 87  
 FT /note= "Mutated amino acid"  
 PN EP-510662-A.  
 PD 28-OCT-1992.  
 PF 24-APR-1992; 107014.  
 PR 26-APR-1991; JP-097655.  
 PR 24-MAR-1992; JP-066381.  
 PA (TAKE ) TAKEDA CHEM IND LTD.  
 PI Fujishima A, Fukuda T;  
 DR WPI; 92-358887/44.  
 DR N-PSDB; Q29698.  
 PT Basic fibroblast growth factor mutein - for treatment of burns,  
 PT wounds, thrombosis and arteriosclerosis, and for post-operative  
 PT tissue healing  
 PS Disclosure; Fig 13; 45pp; English.  
 CC The sequences given in R27964-67 are muteins which are derived from

CC basic fibroblast growth factor (bFGF). These muteins have had  
 CC constituent amino acids replaced by other amino acids, pref. Cys.  
 CC These newly introduced cysteine residues form previously nonexistent  
 CC S-S bonds either between themselves or with one of the four Cys  
 CC residues present in the bFGF molecule. These bonds cause security  
 CC and stabilisation of the higher bFGF structure. The mutations in  
 CC the DNA encoding these muteins causing the changes in amino acid  
 CC sequence were introduced by site-directed mutagenesis.  
 SQ Sequence 146 AA;

Query Match 100.0%; Score 12; DB 1; Length 146;  
 Best Local Similarity 25.0%; Pred. No. 7.87e+03;  
 Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 4 PEDGGSGA 11  
 QY 2 PXXXXXXA 9

# RESULT 15

ID R13017 standard; Protein; 152 AA.  
 AC R13017;  
 DT 19-AUG-1991 (first entry)  
 DE Human lymphotoxin antitumour agent.  
 KW Cancer; liposome; dipalmitoylphosphatidylcholine.  
 OS Homo sapiens.  
 PN J03106821-A.  
 PD 07-MAY-1991.  
 PF 20-SEP-1989; 241754.  
 PR 20-SEP-1989; JP-241754.  
 PA (ELED) Denki Kagaku Kogyo KK.  
 DR WPI; 91-175115/24.  
 PT Antitumour agent contg. lymphotoxin stable in blood - comprises  
 PT human lymphotoxin in liposome obtd. from lipid and surfactant.  
 PS Claim 1; Page 119; 6pp; Japanese.  
 CC The lymphotoxin (LT) agent is stable in blood, and increases  
 CC inhibition of tumour metastasis. Liposome carrying it is obtained  
 CC from dipalmitoylphosphatidylcholine and at least one other kind of  
 CC phosphatidylcholine, and allows lower doses to reach the target site,  
 CC reducing side effects.  
 SQ Sequence 152 AA;

Query Match 100.0%; Score 12; DB 1; Length 152;  
 Best Local Similarity 25.0%; Pred. No. 7.87e+03;  
 Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 142 PSTVFFGA 149  
 QY 2 PXXXXXXA 9

Search completed: Sat Apr 15 01:48:41 2000  
 Job time : 36 secs.



CC acids 33 to 72 (which correspond to residues 10-50 of  
CC beta-galactosidase) is especially favoured.  
CC Enzyme donors carrying the recommended substitutions are used in an  
CC assay to determine the amount of a suspected analyte present in a  
CC sample. They are used in association with enzyme acceptors  
CC (comprising the C-terminal sequence of beta-galactosidase) and an  
CC analyte binding protein. There will be alpha-complementation  
CC between the donor and receptor molecules to restore  
CC beta-galactosidase activity when the analyte binds to its binding  
CC partner. It is this beta-galactosidase activity that is assayed.  
CC See also R08012-3, R08341-2 and R08398.  
SQ Sequence 89 AA;

Query Match 100.0%; Score 12; DB 1; Length 89;  
Best Local Similarity 25.0%; Pred. No. 7.87e+03;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 17 PGNIDPRA 24  
|  
QY 2 PXXXXXXA 9

## RESULT 10

ID R23942 standard; Protein; 91 AA.  
AC R23942;  
DT 15-NOV-1992 (first entry)  
DE Plasmid pM575B region.  
KW Polymerase chain reaction; PCR; UTI; Y46B; HindIII; BamHI.  
OS Synthetic.  
FH Key Location/Qualifiers  
FT peptide 1. .21  
FT protein /label= phoA\_signal\_peptide  
FT 22. .91  
FT /label= Polypeptide\_Y46E  
FN EP-486001-A.  
PD 20-MAY-1992.  
PF 13-NOV-1991; 119378.  
PR 13-NOV-1990; JP-306745.  
PA (MOCH ) MOCHIDA PHARM CO LTD.  
PI Kanamori T, Morishita H, Nobuhara M;  
DR WPI: 92-168622/21.  
DR N-PSDB; Q24152.  
PT New polypeptides comprise amino acid sequence of urinary trypsin  
PT inhibitor - are protease inhibitors for treating e.g. ischaemic  
PT heart disease, thrombosis, arthritis, allergy, shock, etc.  
PS Disclosure: Fig 21: 106pp; English.  
CC The sequence given is encoded by a portion of the plasmid pM575B  
CC between the HindIII and BamHI recognition sites and is the  
CC polypeptide of the invention (polypeptide Y46E).  
CC The DNA sequence encoding Y46E was derived by PCR using the primer  
CC sequences given in Q24142 and Q24144.  
CC Polypeptide Y46E is a modified version of a novel polypeptide which  
CC comprises the amino acid sequence that constitutes a portion of  
CC urinary trypsin inhibitor (UTI). This polypeptide has strong  
CC inhibitory activity against proteases such as trypsin, elastase,  
CC plasmin kallikrein and FXa, and can be used in comps. to treat  
CC diseases caused by these enzymes, eg. operative stress, multiple organ  
CC failure, shock, pancreatitis, ischaemic heart disease, nephritis,  
CC hepatic cirrhosis, thrombosis after revascularisation, oedema caused  
CC by increased vascular permeability, adult respiratory distress  
CC syndrome, rheumatoid arthritis, arthritis and/or allergy, disseminated  
CC intravascular coagulation syndrome. It may also be used to prevent  
CC general blood coagulation. The peptide shows no antigenicity against  
CC humans.  
SQ Sequence 91 AA;

Query Match 100.0%; Score 12; DB 1; Length 91;  
Best Local Similarity 25.0%; Pred. No. 7.87e+03;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 17 PVTKATVA 24  
|  
QY 2 PXXXXXXA 9

## RESULT 11

ID R25324 standard; Protein; 100 AA.  
AC R25324;  
DT 18-MAR-1993 (first entry)  
DE Lv region of human rheumatoid factor antibody.  
KW Light chain; variable region; YES8C; arthritis.  
OS Homo sapiens.  
FH Key Location/Qualifiers  
FT region 24. .34  
FT /note= "CDR1"  
FT region 50. .56  
FT /note= "CDR2"  
FT region 92. .94  
FT /note= "CDR3"  
PD J04267889-A.  
PN 24-SEP-1992.  
PF 22-FEB-1991; 048704.  
PR 22-FEB-1991; JP-048704.  
PA (EZAK/) EZAKI K.  
PA (NISR ) NISSUI PHARM CO LTD.  
DR WPI: 92-368404/45.  
DR N-PSDB; Q29766.  
PT Monoclonal human rheumatoid factor - obtd. by prodn. and  
PT secretion of hybridoma obtd. from cell fusion of human bone  
PT marrow derived lymphocyte and P3U1 mouse myeloma cell  
PS Disclosure: Page 5; 7pp; Japanese.  
CC The sequence shown is the variable region of the light chain of  
CC a human monoclonal antibody rheumatoid factor YES8C. The protein may  
CC be isolated from the bone marrow soln. of a rheumatoid arthritis  
CC patient and used to produce hybridomas, allowing prodn. of the  
CC rheumatoid arthritis factor at constant quality in large quantities.  
CC See also R25325.  
SQ Sequence 100 AA;

Query Match 100.0%; Score 12; DB 1; Length 100;  
Best Local Similarity 25.0%; Pred. No. 7.87e+03;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 49 PRLIYGA 56  
|  
QY 2 PXXXXXXA 9

## RESULT 12

ID R23364 standard; Protein; 102 AA.  
AC R23364;  
DT 29-JUL-1992 (first entry)  
DE GroES structural protein.  
KW Heat shock protein; groES gene.  
OS Streptomyces albus.  
PN W09204452-A.  
PD 19-MAR-1992.  
PF 03-SEP-1991; F00701.  
PR 10-SEP-1990; FR-011186.  
PA (INSP ) INST PASTEUR.  
PI Mazodier P, Guglielmi G;  
DR WPI: 92-114358/14.  
DR NSDB; Q22483.  
PT Recombinant DNA contg. heat inducible promoter and heterologous  
PT gene - also vectors, transformed cells and new heat shock  
PT proteins of streptococcus albus  
PS Disclosure: Fig 5; 50pp; French.  
CC The sequence is that of the GroES protein which is encoded by the  
CC structural gene groES. See also Q22477-Q22486.  
SQ Sequence 102 AA;

Query Match 100.0%; Score 12; DB 1; Length 102;  
Best Local Similarity 25.0%; Pred. No. 7.87e+03;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 42 PDEGVVLA 49

KW insect tolerant agent.  
OS Lycopersicon esculentum.  
FH Key Location/Qualifiers  
FT protein 33..69 /note= "p270 mature protein"  
FT US5175095-A.  
PN 29-DEC-1992.  
PD 19-JUL-1989; 382518.  
PF 19-JUL-1989; US-382518.  
PR 17-JUL-1990; US-554195.  
PA (CALJ ) CALGENE INC.  
PI Houck CM, Martineau BM;  
DR WPI; 93-026940/03.  
DR N-PSDB: Q34942.  
PT DNA constructs contg. tomato p2130 transcriptional initiation  
PT region - useful for modulation of endogenous fruit prods. and for  
PT prodn. of exogenous prods.  
PS Disclosure; Fig 4; 18pp; English.  
CC The sequence given was encoded by the complete DNA sequence of cDNA  
CC clone p270. This gene is controlled by an ovary tissue transcription  
CC initiation control region derived from tomatoes. The control region  
CC initiates abundant mRNA in ovaries prepared from unopened flowers, no  
CC detectable mRNA in ripening fruit but shows increased mRNA in response  
CC to tomato leaf wounding. mRNA is found to localise within the inner  
CC core region of the ovary and the outer region of the ovules (the  
CC integuments). The p270 promoter may find application as a wound  
CC inducible promoter. The native activity of the p270 polypeptide in  
CC the tomato ovary is unknown, but in the leaves, the metallocarboxy-  
CC peptidase inhibitor protein may act as a natural insect tolerant  
CC agent.  
SQ Sequence 77 AA;  
Query Match 100.0%; Score 12; DB 1; Length 77;  
Best Local Similarity 25.0%; Pred. No. 7.87e+03;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
Db 67 PYVGRAMA 74  
QY 2 PXXXXXXA 9  
RESULT 7  
ID P60042 standard; Protein; 77 AA.  
AC P60042;  
DT 08-JUL-1991 (first entry)  
DE Sequence encoded by the leader sequence of alkaline phosphatase A  
DE (pho A) modified to provide a Nari site.  
KW Expression system; secretion; heterologous protein.  
FH Key Location/Qualifiers  
FT peptide 1..21  
FT /label= leader  
FT protein 22..77  
FT /note= "N-terminal Arg is labelled 1"  
PN EP-196864-A.  
PD 08-OCT-1986.  
PF 25-MAR-1986; 302201.  
PR 25-MAR-1985; US-715653.  
PR 07-AUG-1985; US-763932.  
PA (CETU ) CETUS CORP.  
PI Chang S, Lin LSL, Chang SY, Wang AM;  
DR WPI; 86-266619/41.  
DR N-PSDB: N60041.  
PT Transformed prokaryotic cells - with alkaline  
PT phosphatase-mediated processing and secretion of recombinant  
PT proteins  
PS Disclosure; Fig 1; 46pp; English.  
CC The modification of the pho A leader to provide a Nari site permits  
CC coding sequences other than that for pho A to be substituted in  
CC reading frame with leader. However, conversion to the Nari site  
CC prevents processing of the preprotein with respect to alkaline  
CC phosphatase itself since the codon for the N-terminal arginine of  
CC the alkaline phosphatase sequence is thereby converted to that for  
CC proline.

SQ Sequence 77 AA;  
Query Match 100.0%; Score 12; DB 1; Length 77;  
Best Local Similarity 25.0%; Pred. No. 7.87e+03;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
Db 27 PVLENRAA 34  
QY 2 PXXXXXXA 9  
RESULT 8  
ID P70249 standard; Protein; 83 AA.  
AC P70249;  
DT 19-MAY-1991 (first entry)  
DE AA sequence of a polypeptide having human haematopoietic cell  
DE growth potentiating factor (HCGPF) activity.  
KW Autoimmune disease therapy; immunodeficient disease;  
KW bone marrow transplant.  
OS Homo sapiens.  
PN EP-232707-A.  
PD 19-AUG-1987; 100107.  
PF 07-JAN-1987; 100107.  
PR 09-JAN-1986; JP-002633.  
PR 18-DEC-1986; JP-302698.  
PR 08-JAN-1987; JP-002521.  
PA (AJIN ) AJINOMOTO KK.  
PI Tadatsugu T, Gen Y, Junji H, Shinsuke T, Hiroshi M,  
PI Nobukazu K;  
DR WPI; 87-229568/33.  
PT Human haematopoietic cell growth potentiating factor - prepd.  
PT from gene obtd. using RNA from human peripheral blood derived  
PT mononuclear cells  
PS Example; Fig 9; 101pp; English.  
CC The HCGPF exhibits immune control and haematopoietic control functions  
CC over a wide range and may be used in the fields of immunodeficient  
CC diseases, autoimmune diseases, infectious diseases, hepatitis,  
CC nephritis, cancers and bone marrow transplantation.  
SQ Sequence 83 AA;  
Query Match 100.0%; Score 12; DB 1; Length 83;  
Best Local Similarity 25.0%; Pred. No. 7.87e+03;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
Db 58 PYHEEPEA 65  
QY 2 PXXXXXXA 9  
RESULT 9  
ID R08014 standard; protein; 89 AA.  
AC R08014;  
DT 26-FEB-1991 (first entry)  
DE Enzyme donor derived from beta-galactosidase.  
DE Enzyme donor; beta-galactosidase; digoxin immunoassay;  
KW alpha-complementation.  
PN WO9013569-A.  
PD 15-NOV-1990.  
PF 04-MAY-1990; U02491.  
PR 05-MAY-1989; US-347679.  
PA (MICR-) MICROGENICS CORP.  
PI Henderson DR;  
DR WPI; 90-361426/48.  
PT Assay for determination of analyte using enzyme-acceptor and  
PT -donor - detects high mol. wt. proteins e.g. thyroxine, hepatitis  
PT B virus core antigen etc.  
PS Claim 18; Page 93; 101pp; English.  
CC This sequence includes numerous sites where amino acid  
CC substitutions can be made to introduce an amino acid for  
CC conjugation to a ligand. Substitutions can be made at at least one  
CC and not more than two of positions 4, 23, 25, 35, 39, 40, 41,  
CC 42, 43, 44, 45, 48, 52, 55, 61, 68 and 86. Cysteine and lysine are  
CC usually used for the substitutions. The protein consisting of amino

```

PF 17-JAN-1992; U00437.
PR 18-JAN-1991; US-643982.
PI (JOSL-) JOSLIN DIABETES CENT INC.
PA Kahn CR, Rothenberg PL, White MF;
DR WPI; 92-365881/44.
PT Purified nucleic acid encoding Insulin Receptor Substrate - used
PT to prepare IRS-1, for diagnosis and treatment of insulin related
PT diseases and abnormal cellular proliferation
PS Disclosure; Page 25; 128pp; English.
CC The sequences given in R27983-8000 and R28044 are fragments from
CC insulin receptor substrate-1 (IRS-1). These fragments were used to
CC determine the sequence of IRS-1 and to distinguish it from proteins
CC which are co-purified with it. Antibodies were raised against the
CC IRS-1 proteins and were used to remove them from the reaction media.
CC These peptides were formed by proteolytic cleavage of proteins
CC isolated by ID-SPS PAGE to be approx. 185 kd.
SQ Sequence 19 AA;

Query Match 100.0%; Score 12; DB 1; Length 19;
Best Local Similarity 25.0%; Pred. No. 7.87e+03;
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 11 PALTCDEA 18
QY 2 PXXXXXXA 9

RESULT 3
ID P91089 standard; Protein; 47 AA.
AC P91089;
DT 13-MAR-1992 (first entry)
DE Sequence of viper venom polypeptide called "Agkistrostatin".
KW Platelet aggregation inhibitor; antithrombotic agent;
KW myocardial infarction.
OS Viper.
FH Key Location/Qualifiers
FT misc_difference 47
FT /label= OH or at least one AA
PN EP-338634-A.
PD 25-OCT-1989.
PF 17-APR-1989; 200967.
PR 22-APR-1988; US-184653.
PR 22-APR-1988; US-184649.
PR 01-FEB-1989; US-303757.
PA (MERI ) MERCK & CO INC.
PI Friedman PA, Polokoff MA, Gould RJ, Bencen GH, Jacobs JW,
PI Garsky VM, Gan ZB;
DR WPI; 89-311082/43.
PT Viper venom polypeptide cpds. - useful in inhibiting platelet
PT aggregation where strong antithrombotic activity of short
PT duration is needed
PS Claim 4; Page 22; 33pp; English.
CC The polypeptides of the invention have been purified from the venom
CC of various vipers, e.g. Trimeresurus gramineus, E. carinatus,
CC Agkistrodon piscivorus, Bitis arietans and Eristocophis macmahonii.
CC The polypeptides can be used to prevent platelet thrombosis,
CC thromboembolism and reocclusion.
SQ Sequence 47 AA;

Query Match 100.0%; Score 12; DB 1; Length 47;
Best Local Similarity 25.0%; Pred. No. 7.87e+03;
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 3 PANPCDDA 10
QY 2 PXXXXXXA 9

RESULT 4
ID P40034 standard; Protein; 70 AA.
AC P40034;
DT 02-FEB-1992 (first entry)
DE Sequence of human insulin-like growth factor I (IGF-I).

```

```

KW Yeast expression vector; somatic growth; growth promoter.
OS Homo sapiens.
PN EP-123228-A.
PD 31-OCT-1984.
PF 13-APR-1984; 104175.
PR 23-APR-1983; US-487950.
PR 20-SEP-1984; KR-005760.
PA (CHIR-) CHIRON CORP.
PI Barr PJ, Merryweather JP, Mullenbach G, Urdea MS;
DR WPI; 84-271223/44.
DR N-PSDB; N40026.
PT Prodn. of human insulin-like growth factors - by DNA recombinant
PT method, utilising yeast transformant
PS Disclosure; Page 23; 24pp; English.
CC The inventors claim a DNA construct which comprises N40026 or N40027.
CC The DNA constructs are stably replicated in yeasts in which pre-
CC polypeptides form in high yield. The yeast cells are then able to
CC process the pre-forms to the mature IGF.
SQ Sequence 70 AA;

```

```

Query Match 100.0%; Score 12; DB 1; Length 70;
Best Local Similarity 25.0%; Pred. No. 7.87e+03;
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

```

```

Db 63 PLKPAKSA 70
QY 2 PXXXXXXA 9

```

```

RESULT 5
ID P71539 standard; Protein; 70 AA.
AC P71539;
DT 26-MAY-1991 (first entry)
DE Sequence of human insulin-like growth factor I (IGF-I).
KW Hormone; growth promoter.
FH Key Location/Qualifiers
FT disulfide_bond 6..47
FT disulfide_bond 18..61
FT disulfide_bond 48..52
PN J62169733-A.
PD 25-JUL-1987.
PF 22-JAN-1986; 011280.
PR 22-JAN-1986; JP-011280.
PA (FUJI ) FUJISAWA PHARM KK.
DR WPI; 87-246982/35.
PT Human insulin-growth factor, which has a new prim. structure - is
PT prepd. by oxidising reduced form IGF-I and treating the obtd.
PT cpds. by eg chromatography, and is used for incorporating
PT thymidine
PS Claim 2; Page 1; 6pp; Japanese.
CC The IGF-I (and its salts) has strong effect for acceleration of
CC thymidine incorporation into animal cells, suggesting that it has
CC strong growth promoting effect. However it has no blood sugar
CC lowering effect.
SQ Sequence 70 AA;

```

```

Query Match 100.0%; Score 12; DB 1; Length 70;
Best Local Similarity 25.0%; Pred. No. 7.87e+03;
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

```

```

Db 63 PLKPAKSA 70
QY 2 PXXXXXXA 9

```

```

RESULT 6
ID R34939 standard; Protein; 77 AA.
AC R34939;
DT 19-MAY-1993 (first entry)
DE p270 polypeptide.
KW cDNA; clone; p270; ovary; tissue; transcription; initiation; control;
KW region; tomato; unopened flower; ripening fruit; leaf; wound; ovule;
KW integument; inducible; promoter; metallocarboxypeptidase; inhibitor;

```

\*\*\*\*\*

MAISEH (TM)

\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.

Copyright (c) 1993-1998 University of Edinburgh, U.K.

Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Sat Apr 15 01:48:05 2000; MasPar time 3.04 Seconds

Tabular output not generated. 70.164 Million cell updates/sec

Title: >US-08-452-843-26  
Description: (1-9) from US08452843.pep  
Perfect Score: 12  
Sequence: 1 XPXXXXXXA 9

Scoring table: PAM 150  
Gap 15

Searched: 188963 seqs, 23686106 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: a-geneseq  
1: geneseq

Statistics: Mean 6.626; Variance 8.939; scale 0.741

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	12	100.0	17	1	R28071 HSV-1 epitope-contg. m	7.87e+03
2	12	100.0	19	1	R27992 Tryptic peptide fragme	7.87e+03
3	12	100.0	47	1	P91089 Sequence of viper veno	7.87e+03
4	12	100.0	70	1	P40034 Sequence of human insu	7.87e+03
5	12	100.0	70	1	P71539 Sequence of human insu	7.87e+03
6	12	100.0	77	1	R34939 p270 polypeptide.	7.87e+03
7	12	100.0	77	1	P60042 Sequence encoded by th	7.87e+03
8	12	100.0	83	1	P70249 AA sequence of a poly	7.87e+03
9	12	100.0	89	1	R08014 Enzyme donor derived f	7.87e+03
10	12	100.0	91	1	R23942 Plasmid pm575B region.	7.87e+03
11	12	100.0	100	1	R25324 Lv region of human rbe	7.87e+03
12	12	100.0	102	1	R22364 GroES structural prote	7.87e+03
13	12	100.0	107	1	R11581 Macrocyclic FK-506 rece	7.87e+03
14	12	100.0	146	1	R27966 DFGF mutain BFMA.	7.87e+03
15	12	100.0	152	1	R13017 Human lymphotoxin anti	7.87e+03
16	12	100.0	168	1	R42056 Barley Subtilisin inh	7.87e+03
17	12	100.0	174	1	R15619 HBSAg pre-S region sub	7.87e+03
18	12	100.0	198	1	P80603 Sequence of polypeptid	7.87e+03
19	12	100.0	233	1	R12363 HTIV-1 env B antigenic	7.87e+03
20	12	100.0	275	1	R10210 Mutant subtilisin poly	7.87e+03
21	12	100.0	321	1	R81172 Sequence encoded by a	7.87e+03
22	12	100.0	353	1	R20178 P. glumae PGI lipase st	7.87e+03
23	12	100.0	356	1	P71677 aroF gene product from	7.87e+03

24	12	100.0	358	1	R05123 Bat-PA(L).	7.87e+03
25	12	100.0	363	1	R08010 Protein capable of com	7.87e+03
26	12	100.0	375	1	R32009 Rp40-TIA-1.	7.87e+03
27	12	100.0	380	1	P20038 Pre-prorennin-A protei	7.87e+03
28	12	100.0	395	1	R05125 Modified Bat-PA(H).	7.87e+03
29	12	100.0	396	1	P71678 aspc gene product from	7.87e+03
30	12	100.0	447	1	P70314 Sequence of flagellin	7.87e+03
31	12	100.0	457	1	P93628 Sequence of human inte	7.87e+03
32	12	100.0	464	1	R42921 Human antithrombin III	7.87e+03
33	12	100.0	464	1	R42922 Human antithrombin III	7.87e+03
34	12	100.0	609	1	P60064 Recombinant human seru	7.87e+03
35	12	100.0	630	1	R11490 Tissue-plastin.	7.87e+03
36	12	100.0	723	1	R25677 Recombinant human hepa	7.87e+03
37	12	100.0	849	1	R14925 Mutant SP6DNA polymera	7.87e+03
38	12	100.0	858	1	P81779 Sequence encoded by op	7.87e+03
39	12	100.0	914	1	R15785 B.thuringiensis toxin/	7.87e+03
40	12	100.0	934	1	R15048 Soluble human IGF-I re	7.87e+03
41	12	100.0	956	1	R15784 B.thuringiensis toxin/	7.87e+03
42	12	100.0	982	1	R13320 Murine Natural Killer	7.87e+03
43	12	100.0	1148	1	R43671 M.leprae rpoB gene pro	7.87e+03
44	12	100.0	1165	1	R10192 Insecticidal crystal p	7.87e+03
45	12	100.0	1425	1	P80267 Modified factor VIII:C	7.87e+03

#### ALIGNMENTS

RESULT 1  
ID R28071 standard; Protein; 17 AA.  
AC R28071;  
DT 19-MAR-1993 (first entry)  
DE HSV-1 epitope-contg. monomeric peptide.  
KW Herpes Simplex Virus; ELISA; control peptide;  
KW enzyme-linked immunosorbent assay.  
OS Synthetic.  
PN WO9218528-A.  
PD 29-OCT-1992.  
PF 09-APR-1992; G00632.  
PR 09-APR-1991; GB-007434.  
PA (MEDI-) MEDICAL RES COUNCIL.  
PI Marsden H. Subak-Sharpe JH;  
DR WPI; 92-382041/46.  
PT Branded peptides useful in assays for antibodies - contain spacer arms between epitope-contg. portion and polyfunctional core, allowing detection of antibodies at much lower concentrations  
PS Example 1: Page 16; 28pp; English.  
CC The monomeric peptide corresponds to amino acids 357 to 373 of HSV-1 UL42 and was used as a control peptide in an ELISA to test for reactivity with MAB 21F11 which is known to recognise the sequence CC GPEDLD (i.e. amino acids 360-366 of HSV-1). MAB 21F11 was CC significantly more reactive with branched peptides containing 8 CC copies of GPEDLD (see R28069) than with the control monomeric CC peptide.  
SQ Sequence 17 AA;  
Query Match 100.0%; Score 12; DB 1; Length 17;  
Best Local Similarity 25.0%; Pred. No. 7.87e+03;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
Db 6 PEDLDGAA 13  
QY 2 PXXXXXXA 9  
RESULT 2  
ID R27992 standard; peptide; 19 AA.  
AC R27992;  
DT 17-MAR-1993 (first entry)  
DE Tryptic peptide fragment #10.  
KW Insulin receptor substrate-1; IRS-1; antibody; proteolytic cleavage;  
KW ID-SDS PAGE.  
OS Rattus rattus.  
PN WO9213083-A.  
PD 06-AUG-1992.

```
REFERENCE
#authors      JH0092
#journal      Gomez, M.J.; Fluoret, B.; van Heijenoort, J.; Ayala, J.A.
#title        Nucleic Acids Res. (1990) 18:2813
#             Nucleotide sequence of the regulatory region of the gene pbpB
#             of Escherichia coli.
#cross-references MUID:90251464
#accession     JH0092
#molecule_type DNA
#residues      1-346 #label GOM
#cross-references EMBL:X52063; NID:g42317; PID:g42318
#experimental_source strain K-12, substrain W3110
REFERENCE
#authors      S40531
#             Yura, T.; Mori, H.; Nagai, H.; Nagata, T.; Ishihama, A.;
#             Fujita, N.; Isono, K.; Mizobuchi, K.; Nakata, A.
#submission    submitted to the EMBL Data Library, December 1992
#description    Systematic sequencing of the Escherichia coli genome:
#               analysis of the 0-2.4min region.
#accession     S40593
#molecule_type DNA
#residues      1-346 #label YUR
#cross-references EMBL:D10483; NID:g216434; PID:d1001820; PID:g216497
#experimental_source strain K-12
REFERENCE
#authors      A64720
#             Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.;
#             Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.;
#             Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.;
#             Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao,
#             Y.
#journal        Science (1997) 277:1453-1462
#title          The complete genome sequence of Escherichia coli K-12.
#cross-references MUID:97426617
#accession     B64730
#status        nucleic acid sequence not shown; translation not shown
#molecule_type DNA
#residues      34-346 #label BLAT
#cross-references GB:AE000118; GB:U000096; NID:g1786262; PID:g1786270;
#               UWGP:b0082
#experimental_source strain K-12, substrain MG1655
GENETICS
#gene          yabc
#map_position  2 min
#CLASSIFICATION #superfamily Escherichia coli yabc protein
SUMMARY        #length 346 #molecular-weight 38794 #checksum 858
Query Match    100.0%; Score 12; DB 1; Length 346;
Best Local Similarity 25.0%; Pred. No. 5.99e+03;
Matches        2; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
Db 163 PTRGOSAA 170
|
|
QY 2 PXXXXXXA 9

Search completed: Sat Apr 15 01:49:14 2000
Job time : 16 secs.
```

A.; Braun, M.; Brignell, S.C.; Bron, S.; Bron, S.; Brouillet, S.;  
 Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;  
 Choi, S.K.; Codani, J.J.; Connerton, I.F.; Cummings, N.J.;  
 Daniel, R.A.; Denizot, F.; Devine, K.M.; Duesterhoeft, A.;  
 Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.;  
 Fabret, C.; Ferrari, E.; Foulger, D.; Fritz, C.; Fritze, J.;  
 M.; Futita, Y.; Fuma, S.; Gallizzi, A.; Galleron, N.; Ghim,  
 S.Y.; Glaser, P.; Goffeau, A.; Golightly, E.J.; Grandi, G.;  
 Guiseppi, G.; Guy, B.J.; Haga, K.; Halech, J.; Harwood,  
 C.R.; Henaut, A.; Hilbert, H.; Hoisappel, S.; Hosono, S.;  
 Hulio, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karamata, D.;  
 Kashara, Y.; Klier-Blanchard, M.; Klein, C.; Kobayashi,  
 Y.; Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.;  
 Kurita, K.; Lapidus, A.; Lardinois, S.; Lauber, J.;  
 Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.;  
 Maueel, C.; Medique, C.; Medina, N.; Mellado, R.P.; Mizuno,  
 M.; Mosteti, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly,  
 M.; Ogawa, K.; Ogilwa, A.; Oudega, B.; Park, S.H.; Parro,  
 V.; Pohl, T.M.; Portetelle, D.; Porwollik, S.; Prescott,  
 A.M.; Presecan, E.; Pujic, P.; Purnelle, B.; Rapoport, G.;  
 Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.;  
 Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon, E.;  
 Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.;  
 Sekowska, A.; Seror, S.J.; Serror, P.; Shin, B.S.; Soldo,  
 B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Takahashi, H.;  
 Takemaru, K.; Takeuchi, M.; Tamakoshi, A.; Tanaka, T.;  
 Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.;  
 Vandenbol, M.; Vannier, F.; Vassarotti, A.; Viari, A.;  
 Wambutt, R.; Wedler, E.; Wedler, H.; Weitzenecker, T.;  
 Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto,  
 K.; Yata, K.; Yoshida, K.; Yoshikawa, H.F.; Zumstein, E.;  
 Yoshikawa, H.; Danchin, A.

#journal Nature (1997) 390:249-256  
 #title The complete genome sequence of the Gram-positive bacterium  
 #cross-references MUID:98044033  
 #accession AG9590  
 ##status preliminary; nucleic acid sequence not shown;  
 ##molecule\_type DNA  
 ##residues 1-280 #label KUN  
 ##cross-references GB:499117; GB:AL009126; NID:g2634966; PID:eil83796;  
 ##experimental\_source strain 168

GENETICS  
 #gene aroD  
 #superfamily shikimate dehydrogenase; shikimate dehydrogenase  
 #homology

FEATURE  
 58-258 #domain shikimate dehydrogenase homology #label SKD  
 #length 280 #molecular-weight 30642 #checksum 6123

SUMMARY  
 Query Match 100.0%; Score 12; DB 2; Length 280;  
 Best Local Similarity 25.0%; Pred. No. 5.99e+03;  
 Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 204 PLSLQRAA 211  
 QY 2 PXXXXXXA 9

RESULT 13  
 ENTRY probable shikimate 5-dehydrogenase (EC 1.1.1.25) ydib -  
 TITLE Escherichia coli  
 ORGANISM #formal\_name Escherichia coli  
 DATE 12-Sep-1997 #sequence\_revision 17-Sep-1997 #text\_change  
 01-Feb-1999

ACCESSIONS D64927  
 REFERENCE A64720  
 #authors Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.;  
 Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.;  
 Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.;

Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao,  
 Y.  
 #journal Science (1997) 277:1453-1462  
 #title The complete genome sequence of Escherichia coli K-12.  
 #cross-references MUID:97426617  
 #accession D64927  
 ##status nucleic acid sequence not shown; translation not shown  
 ##molecule\_type DNA  
 ##residues 1-288 #label BLAT  
 ##cross-references GB:AE000264; GB:U00096; NID:g1787978; PID:g1787983;  
 ##experimental\_source strain K-12, substrain MG1655

GENETICS  
 #gene ydib  
 #superfamily shikimate dehydrogenase; shikimate dehydrogenase  
 #homology  
 #nucleotide binding; oxidoreductase; P-loop

KEYWORDS  
 FEATURE 63-271 #domain shikimate dehydrogenase homology #label SKD  
 119-126 #region nucleotide-binding motif A (P-loop)  
 SUMMARY #length 288 #molecular-weight 31228 #checksum 1569

Query Match 100.0%; Score 12; DB 2; Length 288;  
 Best Local Similarity 25.0%; Pred. No. 5.99e+03;  
 Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 83 PAAKLIVA 90  
 QY 2 PXXXXXXA 9

RESULT 14  
 ENTRY quinone oxidoreductase (EC 1.6.5.5) - Pseudomonas aeruginosa  
 TITLE #formal\_name Pseudomonas aeruginosa  
 ORGANISM 06-Jun-1995 #sequence\_revision 21-Jul-1995 #text\_change  
 DATE 08-May-1998

ACCESSIONS S52923  
 REFERENCE S52923  
 #authors Hungerer, C.; Troup, B.; Jahn, D.  
 #submission submitted to the EMBL Data Library, February 1995  
 #description Cloning and regulation of the Pseudomonas aeruginosa hemF  
 gene encoding oxygen-dependent coproporphyrinogen III  
 oxidase.

#accession S52923  
 ##status preliminary  
 ##molecule\_type DNA  
 ##residues 1-325 #label HUN  
 ##cross-references EMBL:X85015; NID:g747872; PID:g695692  
 CLASSIFICATION #superfamily alcohol dehydrogenase; long-chain alcohol  
 dehydrogenase homology

KEYWORDS oxidoreductase  
 FEATURE 26-314 #domain long-chain alcohol dehydrogenase homology #label  
 LADH

SUMMARY #length 325 #molecular-weight 35034 #checksum 251

Query Match 100.0%; Score 12; DB 2; Length 325;  
 Best Local Similarity 25.0%; Pred. No. 5.99e+03;  
 Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 273 PEKLQAMA 280  
 QY 2 PXXXXXXA 9

RESULT 15  
 ENTRY yabc protein - Escherichia coli  
 TITLE #formal\_name Escherichia coli  
 ORGANISM 31-Dec-1990 #sequence\_revision 31-Dec-1990 #text\_change  
 DATE 13-Nov-1998  
 ACCESSIONS JH0092; S40593; B64730

KEYWORDS isomerase; streptomycin biosynthesis  
SUMMARY #length 200 #molecular-weight 21956 #checksum 3181

Query Match 100.0%; Score 12; DB 1; Length 200;  
Best Local Similarity 25.0%; Pred. No. 5.99e+03;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 91 PTYRAWEA 98  
|  
Qy 2 PXXXXXXA 9

RESULT 10  
ENTRY K0NH2C #type fragments  
TITLE opacity protein P.IIC precursor - Neisseria gonorrhoeae (strain JS3) (fragments)  
ALTERNATE\_NAMES outer membrane protein P.IIC  
ORGANISM #formal\_name Neisseria gonorrhoeae strain JS3  
DATE 31-Mar-1992 #sequence\_revision 17-Oct-1997 #text\_change 08-May-1998  
ACCESSIONS S03095; S16360  
REFERENCE S03095  
#authors van der Ley, P.  
#journal Mol. Microbiol. (1988) 2:797-806  
#title Three copies of a single protein II-encoding sequence in the genome of *Neisseria gonorrhoeae* JS3: evidence for gene conversion and gene duplication.

#cross-references MUID:89096501  
#accession S03095  
#molecule\_type DNA  
#residues 1-268 #label VAN  
#cross-references EMBL:X12625  
#experimental\_source strain JS3  
#note 241-Val was also found  
#note expression of opacity proteins is regulated by the number of translated repeat elements C1C1R, which code for part of the signal sequence; the protein can only be synthesized when the number of repeats place the start codon in frame with the rest of the protein

REFERENCE S16360  
Barritt, D.S.; Schwalbe, R.S.; Klapper, D.G.; Cannon, J.G.

Infect. Immun. (1987) 55:2026-2031

Antigenic and structural differences among six proteins II expressed by a single strain of *Neisseria gonorrhoeae*.

#cross-references MUID:87306843

#accession S16360

#status preliminary

#molecule\_type protein

#residues 24-34 #label BAR

GENETICS

#gene PIIC

#superfamily opacity protein

KEYWORDS cell surface component; transmembrane protein

FEATURE

1-10,11-23

#domain signal sequence (fragments) #status predicted

#label SIG\

#product opacity protein P.IIC #status experimental

#label MAT\

#domain transmembrane #status predicted #label TM1\

#domain extracellular #status predicted #label EXT1\

#region semivariable region\

#domain transmembrane #status predicted #label TM2\

#domain extracellular #status predicted #label TM3\

#domain extracellular #status predicted #label EXT2\

#region hypervariable region HV1\

#domain transmembrane #status predicted #label TM4\

#domain extracellular #status predicted #label TM5\

#domain extracellular #status predicted #label EXT3\

#region hypervariable region HV2\

#domain transmembrane #status predicted #label TM6\

#domain transmembrane #status predicted #label TM7\

#domain extracellular #status predicted #label EXT4\

260-268  
SUMMARY #domain transmembrane #status predicted #label TM8  
#length 268 #checksum 4200

Query Match 100.0%; Score 12; DB 1; Length 268;  
Best Local Similarity 25.0%; Pred. No. 5.99e+03;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 32 PYVQADLA 39  
|  
Qy 2 PXXXXXXA 9

RESULT 11  
ENTRY S10532 #type complete  
TITLE interleukin-1 alpha precursor - pig hemopoietin-1; IL-1 alpha  
ALTERNATE\_NAMES  
ORGANISM #formal\_name Sus scrofa domestica #common\_name domestic pig  
DATE 20-Feb-1995 #sequence\_revision 22-Nov-1996 #text\_change 05-Feb-1999

ACCESSIONS S10532  
REFERENCE S10532  
#authors Maliszewski, C.R.; Renshaw, B.R.; Schoenborn, M.A.; Urban, J.F.; Baker, P.E.

#journal Nucleic Acids Res. (1990) 18:4282

#title Porcine IL-1 alpha cDNA nucleotide sequence.

#cross-references MUID:9032454

#accession S10532

#status preliminary

#molecule\_type mRNA

#residues 1-270 #label MAL

#cross-references EMBL:X52731; NID:G1987; PID:G1988

COMMENT Produced by activated macrophages, the IL-1 proteins stimulate thymocyte proliferation by inducing IL-2 release, B-cell maturation and proliferation, and fibroblast growth factor activity.

COMMENT IL-1 proteins are involved in the inflammatory response, being identified as endogenous pyrogen, and are reported to stimulate the release of prostaglandin and collagenase from synovial cells.

COMMENT This protein lacks a conventional signal sequence for protein export. Cleavage of a long N-terminal propeptide occurs with secretion, although uncleaved forms are also released. The uncleaved form of interleukin-1alpha, unlike interleukin-1beta, is fully active.

CLASSIFICATION #superfamily interleukin-1

KEYWORDS cytokine; immunoregulation; inflammation; lipoprotein; lymphokine; macrophage; mitogen; myristylation

FEATURE 113-270

#product interleukin-1 alpha #status predicted #label IL1\

#binding\_site myristate (Lys) (covalent) #status predicted

SUMMARY #length 270 #molecular-weight 30788 #checksum 7016

Query Match 100.0%; Score 12; DB 1; Length 270;

Best Local Similarity 25.0%; Pred. No. 5.99e+03;

Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 148 PSGQYIMA 155  
|  
Qy 2 PXXXXXXA 9

RESULT 12

ENTRY A69590 #type complete

TITLE shikimate 5-dehydrogenase aroD - *Bacillus subtilis*

ORGANISM #formal\_name *Bacillus subtilis*

DATE 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 16-Dec-1998

ACCESSIONS A69590

REFERENCE A69580

#authors Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bertero, M.G.; Bessieres, P.; Bolotin, A.; Borchert, S.; Boriss, R.; Boursier, L.; Brans, J.

```

Best Local Similarity 25.0%; Pred. No. 5.99e+03;
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 32 PFETETRA 39
| |
QY 2 PXXXXXXA 9

RESULT 6
ENTRY B27873 #type complete
TITLE allophycocyanin beta chain - Synechococcus sp. (PCC 6301)
ORGANISM 19-Nov-1988 #sequence_revision 19-Nov-1988 #text_change
DATE 08-Nov-1996
ACCESSIONS B27873
REFERENCE A93127
#authors Hounard, J.; Mazel, D.; Moquet, C.; Bryant, D.A.; Tandeau de
Marsac, N.
#journal Mol. Gen. Genet. (1986) 205:404-410
#title Organization and nucleotide sequence of genes encoding core
components of the phycobilisomes from Synechococcus 6301.
#cross-references MUID:87172294
#accession B27873
##molecule_type DNA
##residues 1-161 #label HOU
##experimental_source PCC 6301
CLASSIFICATION #superfamily phycocyanin
KEYWORDS methylated amino acid; photosynthesis; phycocyanobilin
FEATURE 71
#modified_site N4-methylasparagine (Asn) #status
#predicted\
#binding_site phycocyanobilin (Cys) (covalent) #status
#predicted

SUMMARY 81
#length 161 #molecular-weight 17393 #checksum 7518

Query Match 100.0%; Score 12; DB 2; Length 161;
Best Local Similarity 25.0%; Pred. No. 5.99e+03;
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 122 PIGATVQA 129
| |
QY 2 PXXXXXXA 9

RESULT 7
ENTRY S25306 #type complete
TITLE allophycocyanin beta chain - red alga (Cyanidium
caldarium) chloroplast
ORGANISM #formal_name Chloroplast Cyanidium caldarium
DATE 03-Feb-1994 #sequence_revision 03-Feb-1994 #text_change
08-Sep-1997
ACCESSIONS S25306
REFERENCE S25306
#authors Valentín, K.; Maid, U.; Emich, A.; Zetsche, K.
#journal Plant Mol. Biol. (1992) 20:267-276
#title Organization and expression of a phycobiliprotein gene
cluster from the unicellular red alga Cyanidium caldarium.
#cross-references MUID:93004479
#accession S25306
##molecule_type DNA
##residues 1-163 #label VAL
##cross-references EMBL:X57251; NID:g17969; PID:g17970

GENETICS
#gene apcB'
#genome chloroplast
CLASSIFICATION #superfamily phycocyanin
KEYWORDS chloroplast; methylated amino acid; photosynthesis;
phycocyanobilin
FEATURE 72
#modified_site N4-methylasparagine (Asn) #status
#predicted\
#binding_site phycocyanobilin (Cys) (covalent) #status
#predicted

SUMMARY 82
#length 163 #molecular-weight 18742 #checksum 5823

Query Match 100.0%; Score 12; DB 2; Length 163;
Best Local Similarity 25.0%; Pred. No. 5.99e+03;
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 123 PIGATIRA 130
| |
QY 2 PXXXXXXA 9

RESULT 8
ENTRY CUQH #type complete
TITLE plastocyanin precursor - white campion
ORGANISM #formal_name Silene pratensis, Lychnis alba #common_name
white campion, evening lychnis
DATE 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change
20-Feb-1998
ACCESSIONS A24404
REFERENCE A24404
#authors Sneeke, S.; de Groot, M.; van Binsbergen, J.; Weisbeek, P.
#journal Nature (1985) 317:456-458
#title Sequence of the precursor of the chloroplast thylakoid lumen
protein plastocyanin.
#accession A24404
##molecule_type DNA
##residues 1-165 #label SME
##cross-references GB:X02965
CLASSIFICATION #superfamily plastocyanin
KEYWORDS chloroplast; copper; electron transfer; metalloprotein
FEATURE 1-66
#domain transit peptide (chloroplast) #status predicted
#label TNP\
#product plastocyanin #status predicted #label MAT\
#binding_site copper (His, Cys, His, Met) (type 1)
#status predicted

SUMMARY 8327
#length 165 #molecular-weight 16650 #checksum 8327

Query Match 100.0%; Score 12; DB 1; Length 165;
Best Local Similarity 25.0%; Pred. No. 5.99e+03;
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 13 PSFAGLKA 20
| |
QY 2 PXXXXXXA 9

RESULT 9
ENTRY XUSMEG #type complete
TITLE dtdp-4-dehydrorhamnose 3,5-epimerase (EC 5.1.3.13) -
Streptomyces griseus
ORGANISM #formal_name Streptomyces griseus
DATE 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change
05-Sep-1997
ACCESSIONS S18617
REFERENCE S18617
#authors Pissowotzki, K.; Mansouri, K.; Piepersberg, W.
#journal Mol. Gen. Genet. (1991) 231:113-123
#title Genetics of streptomycin production in Streptomyces griseus:
molecular structure and putative function of genes
strELMB2N.
#cross-references MUID:92092953
#accession S18619
##molecule_type DNA
##residues 1-200 #label PIS
##cross-references EMBL:X62567; NID:g49009; PID:g581676
##note the authors translated the initiation codon GTG for
residue 1 as Val

GENETICS
#gene strM
#start_codon GTG
CLASSIFICATION #superfamily dtdp-4-dehydrorhamnose 3,5-epimerase

```



```

RESULT 2
ENTRY   #type complete
TITLE   amicyanin - Paracoccus denitrificans
ORGANISM #formal_name Paracoccus denitrificans
DATE    25-Feb-1994 #sequence_revision 01-Dec-1995 #text_change
08-Sep-1997

ACCESSIONS S12972
REFERENCE  #journal
#authors   van Spanning, R.J.M.; Wansell, C.W.; Reijnders, W.N.M.;
            Oltmann, L.F.; Stouthamer, A.H.
#journal   FEBS Lett. (1990) 275:217-220
#title     Mutagenesis of the gene encoding amicyanin of Paracoccus
            denitrificans and the resultant effect on methyamine
            oxidation.
#cross-references MUID:91085564
#accession  S12972
#status     preliminary
#molecule_type DNA
##residues  1-131 #label SPA
##cross-references EMBL:X55665; NID:945458; PID:945460
CLASSIFICATION #superfamily plastocyanin
SUMMARY       #length 131 #molecular-weight 13983 #checksum 8737

Query Match 100.0%; Score 12; DB 2; Length 131;
Best Local Similarity 25.0%; Pred. No. 5.99e+03;
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 32 PSSEPFPA 39
|-----|
QY 2 PXXXXXXA 9

RESULT 3
ENTRY   #type complete
TITLE   hemoglobin beta chain - ostrich
ORGANISM #formal_name Struthio camelus #common_name ostrich
DATE    20-Sep-1984 #sequence_revision 20-Sep-1984 #text_change
14-Nov-1997

ACCESSIONS A02443
REFERENCE  #journal
#authors   Oberthur, W.; Voelter, W.; Braunitzer, G.
#journal   Hoppe-Seyler's Z. Physiol. Chem. (1980) 361:969-975
#title     Die Sequenz der Haemoglobine von Streifengans (Anser indicus)
            und Straus (Struthio camelus). Inositolphosphat als
            Modulator Der Evolutionsgeschwindigkeit: die ueberraschende
            Sequenz alpha-63 (E12) Valin.
#cross-references MUID:80247760
#accession  A02443
#molecule_type protein
##residues  1-146 #label OBE
CLASSIFICATION #superfamily globin; globin homology
KEYWORDS      blood; chromoprotein; erythrocyte; heme; iron; oxygen carrier
FEATURE       3-146
#domain globin homology #label GLB\
#binding_site oxygen (His) (distal axial ligand) #status
            predicted\
#length 146 #molecular-weight 16296 #checksum 9676

Query Match 100.0%; Score 12; DB 1; Length 146;
Best Local Similarity 25.0%; Pred. No. 5.99e+03;
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 36 PWTQRFFA 43
|-----|
QY 2 PXXXXXXA 9

RESULT 4
ENTRY   #type complete
TITLE   hemoglobin beta chain - black-headed gull
ORGANISM #formal_name Larus ridibundus #common_name black-headed gull
DATE    31-Mar-1990 #sequence_revision 31-Mar-1990 #text_change
14-Nov-1997

ACCESSIONS S00815
REFERENCE  #journal
#authors   Godovac-Zimmermann, J.; Koesters, J.; Braunitzer, G.;
            Goettenboth, R.
#journal   Biol. Chem. Hoppe-Seyler (1988) 369:341-348
#title     Structural adaptation of bird hemoglobins to high-altitude
            respiration and the primary sequences of black-headed gull
            (Larus ridibundus, Charadriiformes) alpha(A)- and
            beta/beta'-chains.
#cross-references MUID:89000193
#accession  S00815
#molecule_type protein
##residues  1-146 #label GOD
##note      78-Ile was also found
CLASSIFICATION #superfamily globin; globin homology
KEYWORDS      blood; chromoprotein; erythrocyte; heme; heterotetramer;
            iron; oxygen carrier
FEATURE       3-146
#domain globin homology #label GLB\
#binding_site oxygen (His) (distal axial ligand) #status
            predicted\
#length 146 #molecular-weight 16260 #checksum 9875

Query Match 100.0%; Score 12; DB 1; Length 146;
Best Local Similarity 25.0%; Pred. No. 5.99e+03;
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 36 PWTQRFFA 43
|-----|
QY 2 PXXXXXXA 9

RESULT 5
ENTRY   #type complete
TITLE   early E4 17K protein 1 - human adenovirus 2
CONTAINS early E4 10K protein
ORGANISM #formal_name Mastadenovirus h2 #common_name human adenovirus
2
#note     host Homo sapiens (man)
DATE      02-Apr-1982 #sequence_revision 02-Apr-1982 #text_change
12-Apr-1996

ACCESSIONS A03804
REFERENCE  #journal
#authors   Herisse, J.; Rigolet, M.; Dupont de Dinechin, S.; Gallibert,
            F.
#journal   Nucleic Acids Res. (1981) 9:4023-4042
#title     Nucleotide sequence of adenovirus 2 DNA fragment encoding for
            the carboxylic region of the fiber protein and the entire
            E4 region.
#cross-references MUID:82059444
#accession  A03804
#molecule_type DNA
##residues  1-153 #label HER
##note      these probable proteins and the introns in the coding
            regions were assigned by correlating EM data, S1
            digestion studies, and the consensus sequences for
            intron splicing

GENETICS   #map_position 91.8-95.2
            #introns      61/3
CLASSIFICATION #superfamily adenovirus early E4 17K protein
KEYWORDS      early protein
FEATURE       56-153
#product early E4 10K protein #status predicted #label
            TPP
SUMMARY      #length 153 #molecular-weight 17404 #checksum 5366

Query Match 100.0%; Score 12; DB 1; Length 153;

```

\*\*\*\*\*  
W P E F  
(TM)  
\*\*\*\*\*  
Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd  
MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Sat Apr 15 01:48:58 2000; MasPar time 3.04 Seconds  
Tabular output not generated.  
118.507 Million cell updates/sec

Title: >US-08-452-843-26  
Description: (1-9) from US08452843.pep  
Perfect Score: 12  
Sequence: 1 PXXXXXXA 9  
Scoring table: PAM 150  
Gap 15  
Searched: 122810 seqs, 40068593 residues  
Post-processing: Minimum Match 0%  
Listing first 45 summaries  
Database: p1r62  
1:p1r1 2:p1r2 3:p1r3 4:p1r4  
Statistics: Mean 9.242; Variance 3.851; scale 2.400

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	12	100.0	117	1	Ig kappa chain precursor	5.99e+03
2	12	100.0	131	2	amicyanin - Paracoccu	5.99e+03
3	12	100.0	146	1	hemoglobin beta chain	5.99e+03
4	12	100.0	153	1	hemoglobin beta chain	5.99e+03
5	12	100.0	156	1	early E4 17K protein	5.99e+03
6	12	100.0	161	2	allophycocyanin beta	5.99e+03
7	12	100.0	163	2	allophycocyanin beta	5.99e+03
8	12	100.0	165	1	plastocyanin precursor	5.99e+03
9	12	100.0	200	1	gDTP-4-dehydrohannos	5.99e+03
10	12	100.0	268	1	opacity protein P.IIC	5.99e+03
11	12	100.0	270	1	interleukin-1 alpha p	5.99e+03
12	12	100.0	280	2	shikimate 5-dehydroge	5.99e+03
13	12	100.0	288	2	probable shikimate 5-	5.99e+03
14	12	100.0	325	2	quinone oxidoreductas	5.99e+03
15	12	100.0	346	1	yabc protein - Escher	5.99e+03
16	12	100.0	354	1	L-lditol 2-dehydrogen	5.99e+03
17	12	100.0	359	2	homoserine dehydrogen	5.99e+03
18	12	100.0	360	1	translation releasing	5.99e+03
19	12	100.0	375	1	alcohol dehydrogenase	5.99e+03
20	12	100.0	375	1	alcohol dehydrogenase	5.99e+03
21	12	100.0	375	1	monocyte surface glyc	5.99e+03
22	12	100.0	392	1	poliovirus receptor s	5.99e+03
23	12	100.0	397	2	cytochrome P450 mycG	5.99e+03

polyketide synthase ( 5.99e+03  
homoserine dehydrogen 5.99e+03  
histidinol dehydrogen 5.99e+03  
histidinol dehydrogen 5.99e+03  
aspartate transaminas 5.99e+03  
homoserine dehydrogen 5.99e+03  
phosphopyruvate hydra 5.99e+03  
dihydrolipoamide dehy 5.99e+03  
gene 61 protein - hum 5.99e+03  
alkaline exonuclease 5.99e+03  
dihydrolipoamide dehy 5.99e+03  
NADH dehydrogenase (u 5.99e+03  
aldehyde dehydrogenas 5.99e+03  
aldehyde dehydrogenas 5.99e+03  
exonuclease (EC 3.1.1 5.99e+03  
NADH dehydrogenase (u 5.99e+03  
59K transcription act 5.99e+03  
peripentonal hexon-as 5.99e+03  
transketolase (EC 2.2 5.99e+03  
transketolase (EC 2.2 5.99e+03  
noncapsid protein NS1 5.99e+03  
protein kinase C (EC 5.99e+03

ALIGNMENTS

RESULT 1  
ENTRY K2HUGM #type fragment  
TITLE Ig kappa chain precursor V-II region (GM607) - human  
ORGANISM #formal\_name Homo sapiens #common\_name man  
DATE 30-Jun-1987 #sequence\_revision 30-Jun-1987 #text\_change 15-Aug-1997  
ACCESSIONS A01889; B24452  
REFERENCE A01889  
#authors Klobbeck, H.G.; Solomon, A.; Zachau, H.G.  
#journal Nature (1984) 309:73-76  
#title Contribution of human V-kappaII germ-line genes to light-chain diversity.  
#cross-references MUID:84191506  
#accession A01889  
##molecule\_type mRNA  
##residues 1-117 #label KLO  
##note the sequence was determined from the differentiated gene

GENETICS GDB:IGKV2  
#gene GDB:IGKV2  
#map\_position 2p12-2p12  
#cross-references GDB:136265  
#accession A01889

COMPLEX An immunoglobulin heterotetramer subunit consists of two identical light (kappa or lambda) and two identical heavy (alpha, delta, epsilon, gamma, or mu) chains usually stabilized by interchain disulfide bonds. In some cases, such as IgA and IgM, the subunits associate into larger oligomers.  
CLASSIFICATION #superfamily immunoglobulin V region; immunoglobulin homology heterotetramer; immunoglobulin

FEATURE 1-4  
#domain signal sequence (fragment) #status predicted  
#product Ig kappa chain V-II region (GM607) #status predicted #label MAT  
#domain immunoglobulin homology #label IGV  
#disulfide\_bonds #status predicted  
#length 117 #checksum 8818

Query Match 100.0%; Score 12; DB 1; Length 117;  
Best Local Similarity 25.0%; Pred. No. 5.99e+03;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
Db 16 PVTGPGEPA 23  
|  
QY 2 PXXXXXXA 9

**THIS PAGE BLANK (USPTO)**

OS Pyrococcus horikoshii.  
 OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=OT3;  
 RX MEDLINE; 98344137.  
 RA KAWABAYASHI Y., SAWADA M., HORIKAWA H., HAIKAWA Y., HINO Y.,  
 RA YAMAMOTO S., SEKINE M., BABA S., KOSUGI H., HOSUYAMA A., NAGAI Y.,  
 RA SAKAI M., OGURA K., OTUKA R., NAKAZAWA H., TAKAMIYA M., OHFUKU Y.,  
 RA FUNAHASHI T., TANAKA T., KUDOH Y., YAMAZAKI J., KUSHIDA N., OGUCHI A.,  
 RA AOKI K., NAKAMURA Y., ROBB T.F., HORIKOSHI K., MASUCHI Y., SHIZUYA H.,  
 RA KIKUCHI H.;  
 RT "Complete sequence and gene organization of the genome of a hyper-  
 thermophilic archaeobacterium, Pyrococcus horikoshii OT3.";  
 RL DNA Res. 5:55-76(1998).  
 DR EMBL: AP000003; BAA29911.1; -. A48BC82A CRC32;  
 SQ SEQUENCE 162 AA; 18635 MW; A48BC82A CRC32;

Query Match 100.0%; Score 12; DB 1; Length 162;  
 Best Local Similarity 25.0%; Pred. No. 6.40e+03;  
 Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 64 PTEYLLAA 71  
 |  
 Qy 2 PXXXXXXA 9

Search completed: Sat Apr 15 01:52:03 2000  
 Job time : 93 secs.

```

O58547;
AC 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DI 01-JAN-1999 (TrEMBLrel. 09, Last annotation update)
DE 153AA LONG HYPOTHETICAL PROTEIN.
GN PH0817.
OS Pyrococcus horikoshii.
OC Archaea; Euryarchaeota; Thermococcales; Pyrococcaceae; Pyrococcus.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=OT3;
RC MEDLINE; 98344137.
RA KAWARABAYASI Y., SAWADA M., HORIKAWA H., HAIKAWA Y., HINO Y.,
RA YAMAMOTO S., SEKINE M., BABA S., KOSUGI H., HOSOIYAMA A., NAGAI Y.,
RA SAKAI M., OGURA K., OTUKA R., NAKAZAWA H., TAKAMIYA M., OHFUKU Y.,
RA FUNAHASHI T., TANAKA T., KUDOH Y., YAMAZAKI J., KUSHIDA N., OGUCHI A.,
RA AOKI K., NAKAMURA Y., ROBB T.F., HORIKOSHI K., MASUCHI Y., SHIZUYA H.,
RA KIKUCHI H.;
RT "Complete sequence and gene organization of the genome of a hyper-
RT thermophilic archaeobacterium, Pyrococcus horikoshii OT3.";
RL DNA Res. 5:55-76(1998).
DR EMBL: AP000003; BAA29910.1; -
SQ SEQUENCE 153 AA; 17617 MW; C99CD399 CRC32;

Query Match 100.0%; Score 12; DB 1; Length 153;
Best Local Similarity 25.0%; Pred. No. 6.40e+03;
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 52 PILYNFA 59
|
Qy 2 PXXXXXXA 9

RESULT 12
ID Q9YC88 PRELIMINARY; PRT; 157 AA.
AC Q9YE24;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DI 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DE 157AA LONG HYPOTHETICAL 6,7-DIMETHYL-8-RIBITYLLOWAZINE SYNTHASE.
GN AP01366.
OS Aeropyrum pernix.
OC Archaea; Crenarchaeota; Aeropyrum.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=KJ;
RC MEDLINE; 99310339.
RA KAWARABAYASI Y., HINO Y., HORIKAWA H., YAMAZAKI S., HAIKAWA Y.,
RA JIN-NO K., TAKAHASHI M., SEKINE M., BABA S., ANKAI A., KOSUGI H.,
RA HOSOIYAMA A., FUKUI S., NAGAI Y., NISHIJIMA K., NAKAZAWA H.,
RA TAKAMIYA M., MASUDA S., FUNAHASHI T., TANAKA T., KUDOH Y.,
RA YAMAZAKI J., KUSHIDA N., OGUCHI A., AOKI K., KUBOTA K., NAKAMURA Y.,
RA NOMURA N., SAKO Y., KIKUCHI H.;
RT "Complete genome sequence of an aerobic hyper-thermophilic
RT crenarchaeon, Aeropyrum pernix K1.";
RL DNA Res. 6:83-101(1999).
DR EMBL: AP000061; BAA80360.1; -
SQ SEQUENCE 157 AA; 16995 MW; 519097D6 CRC32;

Query Match 100.0%; Score 12; DB 1; Length 157;
Best Local Similarity 25.0%; Pred. No. 6.40e+03;
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 121 PGASPLEA 128
|
Qy 2 PXXXXXXA 9

RESULT 13
ID O54650 PRELIMINARY; PRT; 158 AA.
AC O54650;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DI 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)

```

```

DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE FIMBRIAL SUBUNIT PROTEIN PRECURSOR FIMA.
GN FIMA.
OS Bacteroides nodosus (Dichelobacter nodosus).
OC Bacteria; Proteobacteria; gamma subdivision; Dichelobacteriaceae;
OC Dichelobacter.
RN [1]
RP SEQUENCE FROM N.A.
RA CHIMIRE S.C., EGERTON J.R., DHUNGVEL O.P., JOSHI H.D.;
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF038921; AAB96663.1; -
DR EMBL: AF038920; AAB96661.1; -
DR HSSP; P02974; IAY2.
DR PROSITE; PS00409; PROKAR_NTER_METHYL; 1.
DR PFAM; PF00114; pilin; 1.
KW Signal; Methylation.
FT SIGNAL 1 7 POTENTIAL.
FT CHAIN 8 158 FIMBRIAL SUBUNIT PROTEIN.
FT MOD_RES 8 8 METHYLATION (BY SIMILARITY).
SQ SEQUENCE 158 AA; 16511 MW; E210E90B CRC32;

Query Match 100.0%; Score 12; DB 2; Length 158;
Best Local Similarity 25.0%; Pred. No. 6.40e+03;
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 29 PAYNDYIA 36
|
Qy 2 PXXXXXXA 9

RESULT 14
ID Q9YE24 PRELIMINARY; PRT; 162 AA.
AC Q9YE24;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DI 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DE 162AA LONG HYPOTHETICAL PROTEIN.
GN APE0746.
OS Aeropyrum pernix.
OC Archaea; Crenarchaeota; Aeropyrum.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=KJ;
RC MEDLINE; 99310339.
RA KAWARABAYASI Y., HINO Y., HORIKAWA H., YAMAZAKI S., HAIKAWA Y.,
RA JIN-NO K., TAKAHASHI M., SEKINE M., BABA S., ANKAI A., KOSUGI H.,
RA HOSOIYAMA A., FUKUI S., NAGAI Y., NISHIJIMA K., NAKAZAWA H.,
RA TAKAMIYA M., MASUDA S., FUNAHASHI T., TANAKA T., KUDOH Y.,
RA YAMAZAKI J., KUSHIDA N., OGUCHI A., AOKI K., KUBOTA K., NAKAMURA Y.,
RA NOMURA N., SAKO Y., KIKUCHI H.;
RT "Complete genome sequence of an aerobic hyper-thermophilic
RT crenarchaeon, Aeropyrum pernix K1.";
RL DNA Res. 6:83-101(1999).
DR EMBL: AP000060; BAA9723.1; -
SQ SEQUENCE 162 AA; 17808 MW; 31DB6858 CRC32;

Query Match 100.0%; Score 12; DB 1; Length 162;
Best Local Similarity 25.0%; Pred. No. 6.40e+03;
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 93 PRPQVARA 100
|
Qy 2 PXXXXXXA 9

RESULT 15
ID O58548 PRELIMINARY; PRT; 162 AA.
AC O58548;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DI 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DE 162AA LONG HYPOTHETICAL PROTEIN.
GN PH0818.

```

```
DR EMBL; X87299; CAA60728.1; -
FT NON_TER 1
SQ SEQUENCE 115 AA; 12679 MW; 2DFBF67E CRC32;

Query Match 100.0%; Score 12; DB 2; Length 115;
Best Local Similarity 25.0%; Pred. No. 6.40e+03;
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 15 PITVSWEA 22
QY 2 PXXXXXXA 9

RESULT 7
ID Q9YBWL PRELIMINARY; PRT; 124 AA.
AC Q9YBWL
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
DE 124AA LONG HYPOTHETICAL PROTEIN.
GN APEI489
OS Aeropyrum pernix.
OC Archaea; Crenarchaeota; Aeropyrum.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K1.
RX MEDLINE: 99310339.
RA KAWARABAYASI Y., HINO Y., HORIKAWA H., YAMAZAKI S., HAIKAWA Y.,
RA JIN-NO K., TAKAHASHI M., SEKINE M., BABA S., ANKAI A., KOSUGI H.,
RA HOSAYAMA A., FUKUI S., NAGAI Y., NISHIJIMA K., NAKAZAWA H.,
RA TAKAMIYA M., MASUDA S., FUNAHASHI T., TANAKA T., KUDOH Y.,
RA YAMAZAKI J., KUSHIDA N., OGUCHI A., AOKI K., KUBOTA K., NAKAMURA Y.,
RA NOMURA N., SAKO Y., KIKUCHI H.
RT "Complete genome sequence of an aerobic hyper-thermophilic
RT crenarchaeon, Aeropyrum pernix K1."
RL DNA Res. 6:83-101(1999).
DR EMBL; AP000061; BAA80487.1; -.
SQ SEQUENCE 124 AA; 12942 MW; D938081C CRC32;

Query Match 100.0%; Score 12; DB 1; Length 124;
Best Local Similarity 25.0%; Pred. No. 6.40e+03;
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 82 PRGTYTLA 89
QY 2 PXXXXXXA 9

RESULT 8
ID Q9YBWL PRELIMINARY; PRT; 124 AA.
AC Q9YBWL
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
DE 124AA LONG HYPOTHETICAL PROTEIN.
GN APEI696
OS Aeropyrum pernix.
OC Archaea; Crenarchaeota; Aeropyrum.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K1.
RX MEDLINE: 99310339.
RA KAWARABAYASI Y., HINO Y., HORIKAWA H., YAMAZAKI S., HAIKAWA Y.,
RA JIN-NO K., TAKAHASHI M., SEKINE M., BABA S., ANKAI A., KOSUGI H.,
RA HOSAYAMA A., FUKUI S., NAGAI Y., NISHIJIMA K., NAKAZAWA H.,
RA TAKAMIYA M., MASUDA S., FUNAHASHI T., TANAKA T., KUDOH Y.,
RA YAMAZAKI J., KUSHIDA N., OGUCHI A., AOKI K., KUBOTA K., NAKAMURA Y.,
RA NOMURA N., SAKO Y., KIKUCHI H.
RT "Complete genome sequence of an aerobic hyper-thermophilic
RT crenarchaeon, Aeropyrum pernix K1."
RL DNA Res. 6:83-101(1999).
DR EMBL; AP000062; BAA80697.1; -.
SQ SEQUENCE 124 AA; 12718 MW; 59B6FC6D CRC32;

Query Match 100.0%; Score 12; DB 1; Length 124;
Best Local Similarity 25.0%; Pred. No. 6.40e+03;
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 34 PLLTIYA 41
QY 2 PXXXXXXA 9

RESULT 9
ID O57704 PRELIMINARY; PRT; 135 AA.
AC O57704;
DT 01-AUG-1998 (TREMBlrel. 07, Created)
DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)
DT 01-AUG-1998 (TREMBlrel. 07, Last annotation update)
DE PEPM PROTEIN.
GN PEPM.
OS Acidianus ambivalens (Desulfohalobus ambivalens).
OC Archaea; Crenarchaeota; Sulfolobales; Acidianus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LEI 10;
RA KLETZIN A.;
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ225333; CAA12527.1; -.
KW Plasmid.
SQ SEQUENCE 135 AA; 15711 MW; D3711700 CRC32;

Query Match 100.0%; Score 12; DB 1; Length 135;
Best Local Similarity 25.0%; Pred. No. 6.40e+03;
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 93 PKYEDAA 100
QY 2 PXXXXXXA 9

RESULT 10
ID O68315 PRELIMINARY; PRT; 143 AA.
AC O68315;
DT 01-AUG-1998 (TREMBlrel. 07, Created)
DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
DE ALPHA SUBUNIT OF DINITROGENASE REDUCTASE (FE PROTEIN) (FRAGMENT).
GN NIFH.
OS unidentified nitrogen-fixing bacteria.
OC Bacteria.
RN [1]
RP SEQUENCE FROM N.A.
RA OKUMA M., NODA S., KUDO T.;
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB011882; BAA28417.1; -.
DR HSSP; P00456; 1CP2.
DR PROSITE; PS00692; NIFH_FRXC_2; 1.
DR PROSITE; PS00746; NIFH_FRXC_1; 1.
DR PFAM; PF00142; fer4_NiH; 1.
FT NON_TER 1
FT NON_TER 143
SQ SEQUENCE 143 AA; 15201 MW; 00E9F85B CRC32;

Query Match 100.0%; Score 12; DB 2; Length 143;
Best Local Similarity 25.0%; Pred. No. 6.40e+03;
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 76 PEPGVGCA 83
QY 2 PXXXXXXA 9

RESULT 11
ID O58547 PRELIMINARY; PRT; 153 AA.
```

```
RX MEDLINE; 98049343.
RA KLENK H.-P., CLAYTON R.A., TOMB J.-F., WHITE O., NELSON K.E.,
RA KETCHUM K.A., DODSON R.J., GWINN M., HICKEY E.K., PETERSON J.D.,
RA RICHARDSON D.L., KERLAVAGE A.R., GRAHAM D.E., KYRPIDES N.C.,
RA FLEISCHMANN R.D., QUACKENBUSH J., LEE N.H., SUTTON G.G., GILL S.,
RA KIRKNESS E.F., DOUGHERTY B.A., MCKENNEY K., ADAMS M.D., LOFTUS B.,
RA PETERSON S., REICH C.I., MCNEIL L.K., BADGER J.H., GLODEK A., ZHOU L.,
RA OVERBECK R., GOCAYNE J.D., WEIDMAN J.F., McDONALD L., UTTERBACK T.,
RA COTTON M.D., SPRIGGS T., ARTIACH P., KAINE B.P., SYKES S.M.,
RA SADOW P.W., D'ANDREA K.P., BOWMAN C., FUJII C., GARLAND S.A.,
RA MASON T.M., OLSEN G.J., FRASER C.M., SMITH H.O., WOESE C.R.,
RA VENTER J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
reducing archaeon Archaeoglobus fulgidus."
RL Nature 390:364-370(1997).
DR EMBL; AE001075; AAB90816.1; -.
DR TIGR; AF0416; -.
KW Hypothetical protein.
SQ SEQUENCE 62 AA; 6819 MW; 42F76620 CRC32;

Query Match 100.0%; Score 12; DB 1; Length 62;
Best Local Similarity 25.0%; Pred. No. 6.40e+03;
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 15 PLRSMVEA 22
|
QY 2 PXXXXXXA 9

RESULT 3
ID O73974 PRELIMINARY; PRT; 65 AA.
AC O73974;
RX MEDLINE; 98344137.
RA KAWARABAYASI Y., SAWADA M., HORIKAWA H., HAIKAWA Y., HINO Y.,
RA YAMAMOTO S., SEKINE M., BABA S., KOSUGI H., HOSOIYAMA A., NAGAI Y.,
RA SAKAI M., OGURA K., OTURA R., NAKAZAWA H., TAKAMIYA M., OHFUKU Y.,
RA FUNAHASHI T., TANAKA T., KUDOH Y., YAMAZAKI Y., KUSHIDA N., OGUCHI A.,
RA AOKI K., NAKAMURA Y., ROBB T.F., HORIKOSHI K., MASUCHI Y., SHIZUYA H.,
RA KIKUCHI H.;
RT "Complete sequence and gene organization of the genome of a hyper-
thermophilic archaeobacterium, Pyrococcus horikoshii OT3."
RL DNA Res. 5:55-76(1998).
DR EMBL; AP000002; BAA29551.1; -.
SQ SEQUENCE 65 AA; 7485 MW; B6C9E546 CRC32;

Query Match 100.0%; Score 12; DB 1; Length 65;
Best Local Similarity 25.0%; Pred. No. 6.40e+03;
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 34 PKDEVRLA 41
|
QY 2 PXXXXXXA 9

RESULT 4
ID O68939 PRELIMINARY; PRT; 74 AA.
AC O68939;
RX MEDLINE; 96064397.
RA SCHNIDER U., KEEL C., DEFAGO G., HAAS D.;
RT "Tn5-directed cloning of pqg genes from Pseudomonas fluorescens CHA0:
mutational inactivation of the genes results in overproduction of the
antibiotic pyoluteorin."
RL Appl. Environ. Microbiol. 61:3856-3864(1995).

OC Bacteria; Proteobacteria; alpha subdivision; Rhodospirillaceae;
OC Rhodospirillum.
RN [1]
RP SEQUENCE FROM N.A.
RA LOVELESS T.M., BISHOP P.E.;
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF058778; AAC14325.1; -.
FT NON_TER 1
SQ SEQUENCE 74 AA; 8159 MW; AE7ECF9F CRC32;

Query Match 100.0%; Score 12; DB 2; Length 74;
Best Local Similarity 25.0%; Pred. No. 6.40e+03;
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 28 PQMISDPA 35
|
QY 2 PXXXXXXA 9

RESULT 5
ID Q9YE23 PRELIMINARY; PRT; 114 AA.
AC Q9YE23;
RX MEDLINE; 99310339.
RA KAWARABAYASI Y., HINO Y., HORIKAWA H., YAMAZAKI S., HAIKAWA Y.,
RA JIN-NO K., TAKAHASHI M., SEKINE M., BABA S., ANKAI A., KOSUGI H.,
RA HOSOIYAMA A., FUKUI S., NAGAI Y., NISHIJIMA K., NAKAZAWA H.,
RA TAKAMIYA M., MASUDA S., FUNAHASHI T., TANAKA T., KUDOH Y.,
RA YAMAZAKI J., KUSHIDA N., OGUCHI A., AOKI K., KUBOTA K., NAKAMURA Y.,
RA NOMURA N., SAKO Y., KIKUCHI H.;
RT "Complete genome sequence of an aerobic hyper-thermophilic
Crenarchaeon, Aeropyrum pernix K1."
RL DNA Res. 6:83-101(1999).
DR EMBL; AP000060; BAA79724.1; -.
SQ SEQUENCE 114 AA; 12454 MW; D22C698A CRC32;

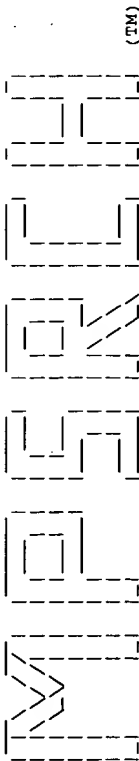
Query Match 100.0%; Score 12; DB 1; Length 114;
Best Local Similarity 25.0%; Pred. No. 6.40e+03;
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 9 PEYSLGAA 16
|
QY 2 PXXXXXXA 9

RESULT 6
ID Q51807 PRELIMINARY; PRT; 115 AA.
AC Q51807;
RX MEDLINE; 96064397.
RA SCHNIDER U., KEEL C., DEFAGO G., HAAS D.;
RT "Tn5-directed cloning of pqg genes from Pseudomonas fluorescens CHA0:
mutational inactivation of the genes results in overproduction of the
antibiotic pyoluteorin."
RL Appl. Environ. Microbiol. 61:3856-3864(1995).

OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonas group;
OC Pseudomonas.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CHA0;
RA SCHNIDER U., KEEL C., DEFAGO G., HAAS D.;
RT "Tn5-directed cloning of pqg genes from Pseudomonas fluorescens CHA0:
mutational inactivation of the genes results in overproduction of the
antibiotic pyoluteorin."
RL Appl. Environ. Microbiol. 61:3856-3864(1995).
```

\*\*\*\*\*



(TM)

\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Sat Apr 15 01:50:30 2000; Maspar time 7.38 Seconds  
Tabular output not generated. 84,503 Million cell updates/sec

Title: >US-08-452-843-26  
Description: (1-9) from US08452843.pep  
Perfect Score: 12  
Sequence: 1 PXXXXXXA 9

Scoring table: PAM 150  
Gap 15

Searched: 225878 seqs, 69334122 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: spiremb112  
1:sp\_archaea 2:sp\_bacteria 3:sp\_fungi 4:sp\_human  
5:sp\_invertebrate 6:sp\_mammal 7:sp\_mhc 8:sp\_organelle  
9:sp\_phase 10:sp\_plant 11:sp\_rodent 12:sp\_unclassified  
13:sp\_vertebrate 14:sp\_virus

Statistics: Mean 9.253; Variance 2.571; scale 3.599

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	12	100.0	36	2	DINITROGENASE 3 BETA S	6.40e+03
2	12	100.0	62	1	CONSERVED HYPOTHETICAL	6.40e+03
3	12	100.0	65	1	65AA LONG HYPOTHETICAL	6.40e+03
4	12	100.0	74	2	DINITROGENASE 3 ALPHA	6.40e+03
5	12	100.0	114	1	114AA LONG HYPOTHETICAL	6.40e+03
6	12	100.0	115	2	UNIDENTIFIED ORF1 (FRA	6.40e+03
7	12	100.0	124	1	124AA LONG HYPOTHETICAL	6.40e+03
8	12	100.0	124	1	124AA LONG HYPOTHETICAL	6.40e+03
9	12	100.0	135	1	PEPM PROTEIN.	6.40e+03
10	12	100.0	143	2	ALPHA SUBUNIT OF DINIT	6.40e+03
11	12	100.0	153	1	153AA LONG HYPOTHETICAL	6.40e+03
12	12	100.0	157	1	157AA LONG HYPOTHETICAL	6.40e+03
13	12	100.0	158	2	FIMBRIAL SUBUNIT PROTE	6.40e+03
14	12	100.0	162	1	162AA LONG HYPOTHETICAL	6.40e+03
15	12	100.0	162	1	162AA LONG HYPOTHETICAL	6.40e+03
16	12	100.0	163	1	163AA LONG HYPOTHETICAL	6.40e+03
17	12	100.0	180	2	ORF20.	6.40e+03
18	12	100.0	182	1	182AA LONG HYPOTHETICAL	6.40e+03
19	12	100.0	193	1	193AA LONG HYPOTHETICAL	6.40e+03
20	12	100.0	231	1	231AA LONG HYPOTHETICAL	6.40e+03

21	12	100.0	254	1	O30288	NUCLEOTIDE-BINDING PRO	6.40e+03
22	12	100.0	257	1	O26936	PROTEIN-EXPORT MEMBRAN	6.40e+03
23	12	100.0	264	1	O28962	CONSERVED HYPOTHETICAL	6.40e+03
24	12	100.0	280	1	O26743	HYPOTHETICAL 31.7 KD P	6.40e+03
25	12	100.0	316	1	O58654	316AA LONG HYPOTHETICA	6.40e+03
26	12	100.0	343	1	O26922	CONSERVED PROTEIN.	6.40e+03
27	12	100.0	363	1	O9YAX6	363AA LONG HYPOTHETICA	6.40e+03
28	12	100.0	365	1	O9Y939	365AA LONG HYPOTHETICA	6.40e+03
29	12	100.0	382	1	O51958	ORF H0026.	6.40e+03
30	12	100.0	390	1	O9Y8Q9	390AA LONG HYPOTHETICA	6.40e+03
31	12	100.0	391	2	O52822	PCZA363.6.	6.40e+03
32	12	100.0	403	1	O26911	CONSERVED PROTEIN.	6.40e+03
33	12	100.0	409	1	O27404	FLAVOPROTEIN AI.	6.40e+03
34	12	100.0	420	2	O52470	FTSA.	6.40e+03
35	12	100.0	429	2	O52702	APALI METHYLTRANSFERAS	6.40e+03
36	12	100.0	431	2	O69275	C2 TOXIN (COMPONENT I)	6.40e+03
37	12	100.0	440	1	O59178	440AA LONG HYPOTHETICA	6.40e+03
38	12	100.0	441	1	O59179	441AA LONG HYPOTHETICA	6.40e+03
39	12	100.0	561	1	O58672	DIHYDROXY-ACID DEHYDRA	6.40e+03
40	12	100.0	585	1	O06504	V-ATPASE A SUBUNIT (EC	6.40e+03
41	12	100.0	780	1	O9YFA5	780AA LONG HYPOTHETICA	6.40e+03
42	12	100.0	902	2	O07686	ORF A PROTEIN.	6.40e+03
43	12	100.0	1097	2	P72196	TONB-LINKED ADHESIN PR	6.40e+03
44	12	100.0	3589	2	O69246	LCHAB PROTEIN.	6.40e+03
45	12	100.0	4735	2	O54666	POLYKETIDE SYNTHASE.	6.40e+03

ALIGNMENTS

RESULT	1	PRELIMINARY;	PRT;	36 AA.
ID	O68941			
AC	O68941			
DC	O1-AUG-1998 (TREMELREL. 07, Created)			
DT	O1-AUG-1998 (TREMELREL. 07, Last sequence update)			
DE	O1-NOV-1999 (TREMELREL. 12, Last annotation update)			
GN	DINITROGENASE 3 BETA SUBUNIT (FRAGMENT).			
OS	ANFK.			
OC	Rhodospirillum rubrum.			
OC	Bacteria; Proteobacteria; alpha subdivision; Rhodospirillaceae;			
OC	Rhodospirillum.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	LOVELESS T.M., BISHOP P.E.;			
RL	Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; AF058778; AAC14327.1;			
DR	HSSP; P11347; IMIO.			
DR	FRAM; PF00148; oxidoredo_nitro; 1.			
FT	NON_TER 36			
SQ	SEQUENCE 36 AA; 3957 MW; 827DE31E CRC32;			

Query Match	100.0%;	Score 12;	DB 2;	Length 36;
Best Local Similarity	25.0%;	Pred. No. 6.40e+03;	Mismatches 0;	Indels 0;
Matches	2;	Conservative 0;	Mismatches 6;	Gaps 0;
Db	16 PFTCOPA 23			
QY	2 PXXXXXXA 9			
RESULT	2			
ID	O29831	PRELIMINARY;	PRT;	62 AA.
AC	O29831			
DC	O1-JAN-1998 (TREMELREL. 05, Created)			
DT	O1-JAN-1998 (TREMELREL. 05, Last sequence update)			
DE	O1-AUG-1998 (TREMELREL. 07, Last annotation update)			
GN	CONSERVED HYPOTHETICAL PROTEIN.			
OS	Archaeoglobus fulgidus.			
OC	Archaea; Euryarchaeota; Archaeoglobales; Archaeoglobaceae;			
OC	Archaeoglobus.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=VC-16 / DSM 4304 / ATCC 49558;			



CC Bacteria; Cyanobacteria; Chroococcales; Synechococcus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 95036003.  
RA BHALLERAO R.P., LIND L.K., GUSTAFSSON P.;  
RT "Cloning of the cpce and cpce genes from Synechococcus sp. PCC 6301  
and their inactivation in Synechococcus sp. PCC 7942.";  
RL Plant Mol. Biol. 26:313-326(1994).  
CC -|- FUNCTION: REQUIRED FOR THE CHROMOPHYLLATION OF THE CPCA GENE  
PRODUCT.  
CC -|- SUBUNIT: CPCE AND CPCE ASSOCIATES TO FORM A LYASE.  
CC -|- SIMILARITY: BELONGS TO THE CPCE/RPCE/PECE FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: M94218; AAA64533.1; -  
KW Phycobilisome; Lyase.  
SQ SEQUENCE 264 AA; 28828 MW; A4340388 CRC32;

Query Match 100.0%; Score 12; DB 1; Length 264;  
Best Local Similarity 25.0%; Pred. No. 4.73e+03;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 169 PDKRLA 176  
|  
QY 2 PXXXXXXA 9

RESULT 14  
ID DCOP\_CANTR STANDARD; PRT: 268 AA.  
AC O42771;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE OROTIDINE 5'-PHOSPHATE DECARBOXYLASE (EC 4.1.1.23) (OMP  
DE DECARBOXYLASE).  
GN URA3.  
OS Candida tropicalis (Yeast).  
OC Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;  
OC Candidaceae; Candida.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-M4;  
RX MEDLINE: 98403409.  
RA SU J.-H., HSIA J.-H., CHANG M.-C.;  
RT "Cloning and sequence analysis of the Candida tropicalis URA3 gene  
RT encoding orotidine-5'-phosphate decarboxylase.";  
RL Curr. Microbiol. 37:210-213(1998).  
CC -|- CATALYTIC ACTIVITY: OROTIDINE-5'-PHOSPHATE -> UMP + CO(2).  
CC -|- PATHWAY: SIXTH AND LAST STEP IN THE BIOSYNTHESIS OF PYRIMIDINES.  
CC -|- SIMILARITY: BELONGS TO THE OMP DECARBOXYLASE FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: AF040702; AAB96773.1; -  
DR PROSITE: PS00156; OMPDECEASE; 1.  
DR PFAM: PF00215; OMPdecase; 1.  
DR Pyrimidine biosynthesis; Lyase; Decarboxylase.  
ACT\_SITE 94 94 BY SIMILARITY.  
SEQUENCE 268 AA; 29672 MW; 3489F3A0 CRC32;

Query Match 100.0%; Score 12; DB 1; Length 268;  
Best Local Similarity 25.0%; Pred. No. 4.73e+03;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 145 PRGLMLA 152  
|  
QY 2 PXXXXXXA 9

RESULT 15  
ID CRTB\_AGRAU STANDARD; PRT: 301 AA.  
AC P54975;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE PHYTOENE SYNTHASE (EC 2.5.1.-).  
GN CRTB.  
OS Agrobacterium aurantiacum.  
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
OC Rhizobiaceae; Agrobacterium.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 96052243.  
RA MISAWA N., SATOMI Y., KONDO K., YOKOYAMA A., KAJIWARA S., SAITO T.,  
RA OHTANI T., MIKI W.;  
RT "Structure and functional analysis of a marine bacterial carotenoid  
RT biosynthesis gene cluster and astaxanthin biosynthetic pathway  
RT proposed at the gene level.";  
RL J. Bacteriol. 177:6575-6584(1995).  
CC -|- FUNCTION: CATALYSES THE REACTION FROM PREPHYTOENE DIPHOSPHATE  
CC TO PHYTOENE.  
CC -|- CATALYTIC ACTIVITY: 2 GERANYLGERANYL DIPHOSPHATE -> PYROPHOSPHATE +  
CC PREPHYTOENE DIPHOSPHATE.  
CC -|- CATALYTIC ACTIVITY: PREPHYTOENE DIPHOSPHATE -> PYROPHOSPHATE +  
CC PHYTOENE.  
CC -|- PATHWAY: CAROTENOID BIOSYNTHESIS. INVOLVED IN ASTAXANTHIN  
CC BIOSYNTHETIC PATHWAY.  
CC -|- SIMILARITY: BELONGS TO THE PHYTOENE/SQUALENE SYNTHETASE FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: D58420; BAA09595.1; -  
DR PROSITE: PS01044; SQUALEN\_PHYTOEN\_SYN\_1; 1.  
DR PROSITE: PS01045; SQUALEN\_PHYTOEN\_SYN\_2; 1.  
DR PFAM: PF00494; SQS\_PSY; 1.  
KW Multifunctional enzyme; Carotenoid biosynthesis; Transferase.  
SQ SEQUENCE 301 AA; 32697 MW; 258DE079 CRC32;

Query Match 100.0%; Score 12; DB 1; Length 301;  
Best Local Similarity 25.0%; Pred. No. 4.73e+03;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 86 PPFACAVA 93  
|  
QY 2 PXXXXXXA 9

Search completed: Sat Apr 15 01:50:12 2000  
Job time : 41 secs.

RL Mol. Genet. 216:254-268(1989).  
CC -1- PATHWAY: CAROTENOID AND CHLOROPHYLL BIOSYNTHESIS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; Z11165; CAA77539.1; -  
DR EMBL; X52291; CAA36532.1; -  
DR PIR; S04401; S04401.  
DR PIR; S17822; S17822.  
KW Photosynthesis; Chlorophyll biosynthesis; Carotenoid biosynthesis;  
KW Oxidoreductase.  
SQ SEQUENCE 241 AA; 27004 MW; 59085F33 CRC32;  
Query Match 100.0%; Score 12; DB 1; Length 241;  
Best Local Similarity 25.0%; Pred. No. 4.73e+03;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
Db -80 PVKRWRA 87  
QY 2 PXXXXXXA 9  
RESULT 11  
ID ADH\_DRODI STANDARD; PRT; 253 AA.  
AC P22245;  
DT 01-AUG-1991 (Rel. 19, Created)  
DT 01-MAY-1992 (Rel. 22, Last sequence update)  
DT 01-OCT-1996 (Rel. 34, Last annotation update)  
DE ALCOHOL DEHYDROGENASE (EC 1.1.1.1).  
GN ADH.  
OS Drosophila differens (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 91163323.  
RA ROWAN R.G., HUNT J.A.;  
RT "Rates of DNA change and phylogeny from the DNA sequences of the  
RT alcohol dehydrogenase gene for five closely related species of  
RT Hawaiian Drosophila."  
RL Mol. Biol. Evol. 8:49-70(1991).  
CC -1- CATALYTIC ACTIVITY: ALCOHOL + NAD(+) = ALDEHYDE OR KETONE + NADH.  
CC -1- SUBUNIT: HOMODIMER.  
CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES  
CC FAMILY (SDR).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; M63303; AAA28350.1; -  
DR FLYBASE; FBgn0012249; ddiif\adh.  
DR PROSITE; PS00061; ADH\_SHORT; 1.  
DR PFAM; PF00106; adh\_short; 1.  
DR PFAM; PF00663; adh\_short\_C; 1.  
KW Oxidoreductase; NAD.  
FT INIT\_MET 0 0 BY SIMILARITY.  
FT NP\_BIND 9 32 NAD (BY SIMILARITY).  
FT ACT\_SITE 150 150 BY SIMILARITY.  
SQ SEQUENCE 253 AA; 27320 MW; B220785F CRC32;  
Query Match 100.0%; Score 12; DB 1; Length 253;

Best Local Similarity 25.0%; Pred. No. 4.73e+03;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
Db 148 PVYSASKA 155  
QY 2 PXXXXXXA 9  
RESULT 12  
ID ADH\_DROER STANDARD; PRT; 255 AA.  
AC P28483;  
DT 01-DEC-1992 (Rel. 24, Created)  
DT 01-DEC-1992 (Rel. 24, Last sequence update)  
DT 01-OCT-1996 (Rel. 34, Last annotation update)  
DE ALCOHOL DEHYDROGENASE (EC 1.1.1.1).  
GN ADH.  
OS Drosophila erecta (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN-GIF-SUR-IVETTE STOCK 154.1;  
RX MEDLINE; 94224125.  
RA JEFFS P.S., HOLMES E.C., ASHBURNER M.;  
RT "The molecular evolution of the alcohol dehydrogenase and alcohol  
RT dehydrogenase-related genes in the Drosophila melanogaster species  
RT subgroup."  
RL Mol. Biol. Evol. 11:287-304(1994).  
CC -1- CATALYTIC ACTIVITY: ALCOHOL + NAD(+) = ALDEHYDE OR KETONE + NADH.  
CC -1- SUBUNIT: HOMODIMER.  
CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES  
CC FAMILY (SDR).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; X54116; CAA38057.1; -  
DR PIR; S20713; S20713.  
DR FLYBASE; FBgn0012261; Dere\adh.  
DR PROSITE; PS00061; ADH\_SHORT; 1.  
DR PFAM; PF00106; adh\_short; 1.  
DR PFAM; PF00663; adh\_short\_C; 1.  
KW Oxidoreductase; NAD.  
FT INIT\_MET 0 0 BY SIMILARITY.  
FT NP\_BIND 11 34 NAD (BY SIMILARITY).  
FT ACT\_SITE 152 152 BY SIMILARITY.  
SQ SEQUENCE 255 AA; 27592 MW; 2C2121CA CRC32;  
Query Match 100.0%; Score 12; DB 1; Length 255;  
Best Local Similarity 25.0%; Pred. No. 4.73e+03;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
Db 150 PVYSGTKA 157  
QY 2 PXXXXXXA 9  
RESULT 13  
ID CPCE\_SYNP7 STANDARD; PRT; 264 AA.  
AC Q44115;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE PHYCOCYANOBILIN LYASE ALPHA SUBUNIT (EC 4.1.1.1) (PHYCOCYANIN OPERON  
DE PROTEIN CPCE).  
GN CPCE.  
OS Synechococcus sp. (strain PCC 7942) (Anacystis nidulans R2).

```
FT MOD_RES 1 1 ACERYLATION (PROBABLE).
SQ SEQUENCE 173 AA; 19769 MW; D44E754A CRC32;

Query Match
Best Local Similarity 25.0%; Score 12; DB 1; Length 173;
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 121 PSSVDQSA 128
QY 2 PXXXXXXA 9

RESULT 8
ID CAC2_HAECO STANDARD; PRT; 210 AA.
AC P16252;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 01-MAY-1992 (Rel. 22, Last annotation update)
DE CUTICLE COLLAGEN 2C (FRAGMENT).
GN 2C.
OS Haemochus contortus.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Strongylida;
OC Trichostrongyloidea; Trichostrongylidae; Haemonchidae; Haemonchus.
RN [1]
RP SEQUENCE FROM N.A.
RX SHAMANSKY L.M., PRATT D., BOISVENUE R.J., COX G.N.;
RT "Cuticle collagen genes of Haemonchus contortus and Caenorhabditis
RT elegans are highly conserved."
RL Mol. Biochem. Parasitol. 37:73-86(1989).
CC -!- FUNCTION: NEMATODE CUTICLES ARE COMPOSED LARGELY OF COLLAGEN-LIKE
CC PROTEINS. THE CUTICLE FUNCTIONS BOTH AS AN EXOSKELETON AND AS A
CC BARRIER TO PROTECT THE WORM FROM ITS ENVIRONMENT.
CC -!- MISCELLANEOUS: THIS PROTEIN SHOWS 4 POTENTIAL TRIPLE-HELICAL
CC REGIONS, WHICH CONTAIN GLYCINE AS EVERY THIRD AMINO ACID.
CC -!- MISCELLANEOUS: IN ALL NEMATODE CUTICLE COLLAGENS, THE POLYPEPTIDE
CC CHAINS ARE COMPLEXED WITHIN THE CUTICLE BY DISULFIDE BONDS AND
CC OTHER TYPES OF COVALENT CROSS-LINKS.
CC -!- SIMILARITY: TO CAENORHABDITIS ELEGANS CUTICLE COLLAGENS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; J04670; AAA29172.1;
CC PIR; P01391; Collagen; 2;
CC DR PFAM; PF01391; Collagen; 2;
CC KW Cuticle; Connective tissue; Repeat; Multigene family; Collagen.
CC FT NON_TER 1
CC SQ SEQUENCE 210 AA; 19562 MW; C660A7EB CRC32;

Query Match
Best Local Similarity 25.0%; Score 12; DB 1; Length 210;
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 82 PGAPGND 89
QY 2 PXXXXXXA 9

RESULT 9
ID CREA_HUMAN STANDARD; PRT; 220 AA.
AC Q03060;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 15-DEC-1999 (Rel. 39, Last annotation update)
DE CAMP-RESPONSIVE ELEMENT MODULATOR, ALPHA ISOFORM.
GN CREM.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Hominidae; Homo.
[1]
RN SEQUENCE FROM N.A.
RC TISSUE-PLACENTA;
RX MEDLINE; 93096608.
RA MEYER T.E., HABENER J.F.;
RT "Cyclic AMP response element binding protein CREB and modulator
RT protein CREM are products of distinct genes."
RL Nucleic Acids Res. 20:6106-6106(1992).
CC -!- FUNCTION: THIS PROTEIN BINDS THE CAMP RESPONSE ELEMENT (CRE),
CC A SEQUENCE PRESENT IN MANY VIRAL AND CELLULAR PROMOTERS. CREM
CC ALPHA, BETA, AND GAMMA ISOFORMS ARE ANTAGONISTS OF THE CAMP
CC TRANSCRIPTIONAL RESPONSE, WHILE THE DELTA ISOFORM IS AN ACTIVATOR.
CC -!- SUBUNIT: BINDS DNA AS A DIMER (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- ALTERNATIVE PRODUCTS: ALPHA, BETA, GAMMA, AND DELTA ISOFORMS OF
CC CREM ARE PRODUCED BY ALTERNATIVE SPLICING OF A SINGLE GENE.
CC -!- PTM: STIMULATED BY PHOSPHORYLATION (BY SIMILARITY).
CC -!- SIMILARITY: TO OTHER BZIP PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z15159; CAA78858.1;
CC PIR; S26885; S26885.
CC DR HSP; P03412; IFOS.
CC DR TRANSFAC; T01803;
CC DR MIM; 123812;
CC DR PROSITE; PS00036; BZIP_BASIC; 1.
CC DR PFAM; PF00170; bZIP; 1.
CC KW Transcription regulation; DNA-binding; Repressor; Phosphorylation;
CC Nuclear protein; Alternative splicing.
CC FT DNA_BIND 163 184
CC FT DOMAIN 190 211
CC FT LEUCINE-ZIPPER (BY SIMILARITY).
CC SQ SEQUENCE 220 AA; 24307 MW; 76607C4C CRC32;

Query Match
Best Local Similarity 25.0%; Score 12; DB 1; Length 220;
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 36 PALRQVAA 43
QY 2 PXXXXXXA 9

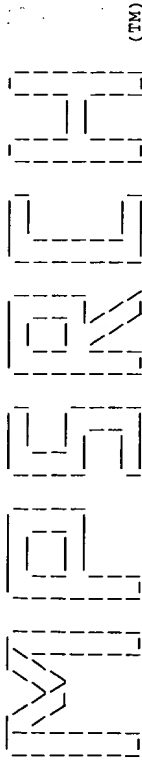
RESULT 10
ID CRTA_RHOCA STANDARD; PRT; 241 AA.
AC P17055;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 01-MAY-1992 (Rel. 22, Last annotation update)
DE SPHEROIDE MONOOXYGENASE (EC 1.-.-.-).
GN CRTA.
OS Rhodobacter capsulatus (Rhodospseudomonas capsulata).
OC Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;
OC Rhodobacter.
RN [1]
RP SEQUENCE FROM N.A.
RA BURKE D.H., ALBERTI M., ARMSTRONG G.A., HEARST J.E.;
RL Submitted (NOV-1991) to the EMBL/GenBank/DBJ databases.
RN [2]
RP PRELIMINARY SEQUENCE FROM N.A.
RC STRAIN-SB1003, AND BEC404;
RX MEDLINE; 89313663.
RA ARMSTRONG G.A., ALBERTI M., LEACH F., HEARST J.E.;
RT "Nucleotide sequence, organization, and nature of the protein
RT products of the carotenoid biosynthesis gene cluster of Rhodobacter
RT capsulatus.";
```



OC Pterygota; Neoptera; Orthopteroidea; Orthoptera; Caelifera;  
OC Acridomorpha; Acridoidea; Acrididae; Oedipodinae; Locusta.  
RN [1]  
RP SEQUENCE.  
RX MEDLINE: 94039045.  
RA ANDREASEN L., HOEJURUP P., ANDERSEN S.O., ROEPSTORFF P.;  
RT "Combined plasma-desorption mass spectrometry and Edman degradation  
RT applied to simultaneous sequence determination of isoforms of  
RT structural proteins from the cuticle of Locusta migratoria";  
RL Eur. J. Biochem. 217:267-273(1993).  
CC -|- FUNCTION: COMPONENT OF THE CUTICLE OF MIGRATORY LOCUST WHICH  
CC CONTAINS MORE THAN 100 DIFFERENT STRUCTURAL PROTEINS.  
CC -|- DOMAIN: THE TRAPEZITIDE (A-A-P-[AV]) REPEATS FOUND THROUGHOUT THE  
CC PROTEIN ARE ALSO PRESENT IN MANY PROTEINS CONSTITUTING THE  
CC PROTECTIVE ENVELOPE OF OTHER SPECIES.  
CC -|- MISCELLANEOUS: THE SEQUENCE SHOWN HERE IS THAT OF ISOFORM LM-70A.  
DR PIR: S38267; S38267. Cuticle; Repeat.  
KW Structural protein; Cuticle; Repeat.  
FT REPEAT 7 10 1.  
FT REPEAT 48 51 2.  
FT REPEAT 55 58 3.  
FT REPEAT 60 63 4.  
FT REPEAT 66 69 5.  
FT VARIAT 82 82 F -> Y (IN ISOFORM LM-70B).  
SQ SEQUENCE 88 AA; 8314 MW; E5CC0491 CRC32;  
  
Query Match 100.0%; Score 12; DB 1; Length 88;  
Best Local Similarity 25.0%; Pred. No. 4.73e+03;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
  
Db 9 PAVAVAPA 16  
QY 2 PXXXXXXA 9  
  
RESULT 3  
ID DEF1-ANOGA STANDARD; PRT; 102 AA.  
AC Q17027;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE DEFENSIN PRECURSOR.  
OS Anopheles gambiae (African malaria mosquito).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;  
OC Culicidae; Anopheles.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-G3;  
RA RICHMANN A.M.;  
RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.  
CC -|- FUNCTION: RESPONSIBLE FOR THE ANTI GRAM-POSITIVE ACTIVITY OF  
CC IMMUNE HEMOLYMPH (BY SIMILARITY).  
CC -|- SIMILARITY: BELONGS TO THE ARTHROPOD DEFENSIN FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
DR EMBL: X93562; CAA63775.1; ALT\_INIT.  
DR HSP; P10891; LICA.  
DR PROSITE; PS00425; ARTHROPOD\_DEFENSINS; 1.  
DR PFAM; PF01097; Defensin; 1.  
KW Insect immunity; Antibiotic; Signal.  
FT SIGNAL 1 25  
FT PROPEP 26 62 POTENTIAL.  
FT CHAIN 63 102 DEFENSIN.  
FT -DISULFID 65 92 BY SIMILARITY.  
FT -DISULFID 78 98  
BY SIMILARITY.

FT DISULFID 82 100 BY SIMILARITY.  
SQ SEQUENCE 102 AA; 10627 MW; 6CC89F1F CRC32;  
  
Query Match 100.0%; Score 12; DB 1; Length 102;  
Best Local Similarity 25.0%; Pred. No. 4.73e+03;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
  
Db 47 PEETHAA 54  
QY 2 PXXXXXXA 9  
  
RESULT 4  
ID CHPA\_ECOLI STANDARD; PRT; 111 AA.  
AC P33645;  
DT 01-FEB-1994 (Rel. 28, Created)  
DT 01-FEB-1994 (Rel. 28, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE PEMK-LIKE PROTEIN 1 (MAZF PROTEIN).  
GN CHPA OR MAZF OR CHPAK.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Escherichia.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-K12;  
RX MEDLINE: 94042847.  
RA MASUDA Y., MIYAKAWA K., NISHIMURA Y., OHTSUBO E.;  
RT "chpA and chpB, Escherichia coli chromosomal homologs of the pem  
RT locus responsible for stable maintenance of plasmid R100";  
RL J. Bacteriol. 175:6850-6856(1993).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-K12;  
RA AIZENMAN E., GLASER G.;  
RL Submitted (NOV-1993) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-K12 / MG1655;  
RX MEDLINE: 97426617.  
RA BLATNER F.R., PLUNKETT G. III, BLOCH C.A., PERNA N.T., BURLAND V.,  
RA RILEY M., COLLADO-VIDES J., GLASNER F.D., RODE C.K., MAYHEW G.F.,  
RA GREGOR J., DAVIS N.W., KIRKPATRICK H.A., GORDEN M.A., ROSE D.J.,  
RA MAU B., SHAO Y.;  
RT "The complete genome sequence of Escherichia coli K-12";  
RL Science 277:1453-1474(1997).  
CC -|- FUNCTION: MAY BE INVOLVED IN THE REGULATION OF CELL GROWTH. IT ACT  
CC AS A GROWTH INHIBITOR. BOTH CHPR AND CHPA BIND TO THE PROMOTER  
CC REGION OF THE CHPR OPERON TO AUTOREGULATE THEIR SYNTHESIS.  
CC -|- SIMILARITY: BELONGS TO THE PEMK FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
DR EMBL: D16450; BAA03918.1;  
DR EMBL: J04039; AAA03239.1;  
DR EMBL: U29580; AAA69292.1;  
DR EMBL: AE000362; AAC75824.1;  
DR PIR: B49339; B49339.  
DR ECOGENE; EG11249; CHPA.  
KW DNA-binding.  
SQ SEQUENCE 111 AA; 12098 MW; 2D76C4B0 CRC32;  
  
Query Match 100.0%; Score 12; DB 1; Length 111;  
Best Local Similarity 25.0%; Pred. No. 4.73e+03;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
  
Db 19 PTKGSEQA 26

\*\*\*\*\*



\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Sat Apr 15 01:49:31 2000; MasPar time 3.04 Seconds  
Tabular output not generated. 88.479 Million cell updates/sec.

Title: >US-08-452-843-26  
Description: (1-9) from US08452843.pep  
Sequence: 12  
1 XPXXXXXXA 9

Scoring table: PAM 150  
Gap 15

Searched: 82229 seqs, 29864866 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: swiss-prot38  
1:swissprot

Statistics: Mean 9.723; Variance 3.177; scale 3.061

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	12	100.0	61	1	ANK7_BOVIN ANNEXIN VII (SYNEXIN)	4.73e+03
2	12	100.0	88	1	CUTICLE PROTEIN 70, IS	4.73e+03
3	12	100.0	102	1	DEF1_ANGOA DEFENSIN PRECURSOR,	4.73e+03
4	12	100.0	111	1	CHPA_ECOLI PEMK-LIKE PROTEIN 1 (M	4.73e+03
5	12	100.0	115	1	CKN_HUMAN PROCHOLECYSTOKININ PRE	4.73e+03
6	12	100.0	161	1	DNEI_CHLVU DNA ENDONUCLEASE I-CVU	4.73e+03
7	12	100.0	173	1	CHAA_ALLMI ALPHA CRYSTALLIN A CHA	4.73e+03
8	12	100.0	210	1	CAC2_HAECO CUTICLE COLLAGEN 2C (F	4.73e+03
9	12	100.0	220	1	CREA_HUMAN CAMP-RESPONSIVE ELEMEN	4.73e+03
10	12	100.0	241	1	CRTH_RHOCA SPHERIDENE MONOOXYGEN	4.73e+03
11	12	100.0	253	1	ADH_DRODI ALCOHOL DEHYDROGENASE	4.73e+03
12	12	100.0	255	1	ADHL_DROER ALCOHOL DEHYDROGENASE	4.73e+03
13	12	100.0	264	1	CPCE_SYNP7 PHYCOCYANOBILIN LYASE	4.73e+03
14	12	100.0	268	1	DCOP_CANTR OROTIDINE 5'-PHOSPHATE	4.73e+03
15	12	100.0	301	1	CRTB_AGRAU PHYTOENE SYNTHASE (EC	4.73e+03
16	12	100.0	314	1	ARCC_CLOPE CARBAMATE KINASE (EC 2	4.73e+03
17	12	100.0	349	1	CTGF_FIG CONNECTIVE TISSUE GROW	4.73e+03
18	12	100.0	362	1	DCAM_PHANI S-ADENOSYLMETHIONINE D	4.73e+03
19	12	100.0	395	1	DNAJ_MYCTU DNAJ PROTEIN.	4.73e+03
20	12	100.0	424	1	BCHN_RHOCA PROTOCHLOROPHYLLIDE RE	4.73e+03
21	12	100.0	426	1	PROTEIN PHOSPHATASE PP	4.73e+03
22	12	100.0	433	1	DCUA_WOLSU ANAEROBIC C4-DICARBOXY	4.73e+03
23	12	100.0	441	1	DCDA_BACSU DIAMINOPIMELATE DECARB	4.73e+03

24	12	100.0	446	1	DCUB_ECOLI ANAEROBIC C4-DICARBOXY	4.73e+03
25	12	100.0	447	1	2ABA_HUMAN PROTEIN PHOSPHATASE PP	4.73e+03
26	12	100.0	455	1	CC40_YEAST CELL DIVISION CONTROL	4.73e+03
27	12	100.0	461	1	BENA_ACICA BENZOATE 1,2-DIOXYGENA	4.73e+03
28	12	100.0	469	1	AJAC_DIDMA ALPHA-2C ADRENERGIC RE	4.73e+03
29	12	100.0	484	1	1A1C_SOYBN 1-AMINOCYCLOPROPANE-1-	4.73e+03
30	12	100.0	492	1	ATPB_PINTH ATP SYNTHASE BETA CHAI	4.73e+03
31	12	100.0	511	1	C7C4_ARATH CYTOCHROME P450 76C4 (	4.73e+03
32	12	100.0	513	1	ATPA_HARIN ATP SYNTHASE ALPHA CHA	4.73e+03
33	12	100.0	515	1	CF51_PENIT CYTOCHROME P450 51 (EC	4.73e+03
34	12	100.0	516	1	COX1_CYPCA CYTOCHROME C OXIDASE P	4.73e+03
35	12	100.0	526	1	COX1_CYPCA CYTOCHROME C OXIDASE P	4.73e+03
36	12	100.0	538	1	AROF_SOLTU PHOSPHO-2-DEHYDRO-3-DE	4.73e+03
37	12	100.0	546	1	CHOD_STRSQ CHOLESTEROL OXIDASE PR	4.73e+03
38	12	100.0	562	1	APY_AEDAE APYRASE PRECURSOR (EC	4.73e+03
39	12	100.0	578	1	ASO_TOBAC L-ASCORBATE OXIDASE PR	4.73e+03
40	12	100.0	688	1	ANK2_RAT BETA-ADRENERGIC RECEPT	4.73e+03
41	12	100.0	718	1	CDGT_BACLI CYCLOMALTODEXTRIN GLUC	4.73e+03
42	12	100.0	718	1	CDGT_BACLI CYCLOMALTODEXTRIN GLUC	4.73e+03
43	12	100.0	1001	1	ATCA_RABIT CALCIUM-TRANSPORTING A	4.73e+03
44	12	100.0	1048	1	ANGR_VIBAN ANGR PROTEIN.	4.73e+03
45	12	100.0	2201	1	ABCI_MOUSE ATP-BINDING CASSETTE T	4.73e+03

ALIGNMENTS

RESULT	1	STANDARD;	PRT;	61 AA.
AC	P20072;			
DT	01-FEB-1991 (Rel. 17, Created)			
DT	01-FEB-1991 (Rel. 17, Last sequence update)			
DT	01-NOV-1995 (Rel. 32, Last annotation update)			
DE	ANNEXIN VII (SYNEXIN) (FRAGMENT).			
GN	ANK7.			
OS	Bos taurus (Bovine).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;			
OC	Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae;			
OC	Bovinae; Bos.			
RN	[1]			
RP	SEQUENCE.			
RX	MEDLINE; 88240346.			
RA	CREUTZ C.E., SNYDER S.L., HUSTED L.D., BEGGERLY L.K., FOX J.W.;			
RT	"Pattern of repeating aromatic residues in synexin. Similarity to the			
RT	cytoplasmic domain of synaptophysin."			
RL	Biochem. Biophys. Res. Commun. 152:1298-1303(1988).			
CC	-!- FUNCTION: CALCIUM/PHOSPHOLIPID-BINDING PROTEIN WHICH PROMOTES			
CC	MEMBRANE FUSION AND IS INVOLVED IN EXOCYTOSIS.			
CC	-!- SIMILARITY: BELONGS TO THE ANNEXIN FAMILY.			
DR	PIR; A27695; A27695.			
DR	PROSITE; PS00223; ANNEXIN; PARTIAL.			
KW	Annexin; Calcium/phospholipid-binding; Repeat.			
FT	NON_TER			
FT	NON_TER			
SQ	SEQUENCE 61 AA; 5816 MW; AA269AD6 CRC32;			
Query Match 100.0%; Score 12; DB 1; Length 61;				
Best Local Similarity 25.0%; Pred. No. 4.73e+03;				
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;				
Db	26 PSSGYPGA 33			
QY	2 PXXXXXXA 9			
RESULT	2	STANDARD;	PRT;	88 AA.
ID	CU70_LOCOMI			
AC	P80232;			
DT	01-OCT-1993 (Rel. 27, Created)			
DT	01-OCT-1993 (Rel. 27, Last sequence update)			
DT	01-FEB-1996 (Rel. 33, Last annotation update)			
DE	CUTICLE PROTEIN 70, ISOFORMS A AND B (LM-70A AND LM-70B).			
OS	Locusta migratoria (Migratory locust).			
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;			

AC R27934;  
 DT 25-NOV-1992 (first entry)  
 DE GAG fusion protein with SOD according to a formula.  
 KW Glycosamino:glycan; superoxidisedismutase; tissue damage;  
 KW autoimmune disease; rheumatoid arthritis; osteoarthritis; ss.  
 OS Synthetic.  
 FH Key Location/Qualifiers  
 FT peptide 1..172  
 FT peptide /note= "SOD"  
 FT peptide 173..221  
 FT peptide /note= "GAG binding motif"  
 PN W09207935-A.  
 PD 14-MAY-1992.  
 PF 01-NOV-1991; U08105.  
 PR 01-NOV-1990; US-608539.  
 PR 02-NOV-1990; US-608569.  
 PA (SCRI ) SCRIPPS RES INST.  
 PI Boissinot M, Fisher C, Griffin JH, Hallewell RA, Kuhn L;  
 PI Mullenbachgt, Parge HE, Tainer JA;  
 DR WPI; 92-183671/22.  
 RT Fusion proteins with glycosamino:glycan-binding and  
 PT superoxidisedismutase activities - reduce tissue damage caused by  
 PT superoxide radicals, useful in treating autoimmune diseases e.g.  
 PT rheumatoid arthritis and osteoarthritis  
 PS Claim 8; Fig 1; 140pp; English.  
 CC The fusion protein comprising the a glycosaminoglycan binding region  
 CC and human superoxide dismutase, joined via a linker region was  
 CC constructed according to the formula SOD-(M-2)3-M where 2 is the peptide  
 CC -RRHHPREMKRVEDL-. The fusion protein is useful for extending  
 CC the in vivo lifetimes of biologically active epds. such as SOD and  
 CC for targeting them to specific cell surfaces or substrates. The  
 CC glycosaminoglycan (GAG) binding protein is formed into a fusion  
 CC protein with SOD to increase stability, plasma half-life and ease  
 CC of purification of SOD. SOD is useful for reduction of tissue damage  
 CC caused by oxygen radicals and is used in the treatment of autoimmune  
 CC diseases e.g. rheumatoid and osteo-arthritis.  
 CC See also R24225-35, R27932-51.  
 SQ Sequence 221 AA;

Query Match 100.0%; Score 15; DB 1; Length 221;  
 Best Local Similarity 25.0%; Pred. No. 2.97e+03;  
 Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 75 PKDERHV 82  
 Oy 2 PXXXXXV 9

Search completed: Sat Apr 15 01:55:46 2000  
 Job time : 36 secs.

```
FT peptide 1. .32
FT /note= "GAG binding motif"
FT peptide 33. .205
FT /note= "SOD"
PN WO9207935-A.
PD 14-MAY-1992.
PF 01-NOV-1991: U08105.
PR 01-NOV-1990: US-608539.
PR 02-NOV-1990: US-608569.
PA (SCRI ) SCRIPPS RES INST.
PI Boissinot M, Fisher C, Griffin JH, Hallewell RA, Kuhn L;
PI Mullenbachgt, Parge HE, Tainer JA;
DR WPI: 92-183671/22.
PT Fusion proteins with glycosamino:glycan-binding and
PT superoxidismutase activities - reduce tissue damage caused by
PT super:oxide radicals, useful in treating autoimmune diseases e.g.
PT rheumatoid arthritis and osteoarthritis
PS Claim 8; Fig 1: 140pp; English.
CC The fusion protein comprising the a glycosaminoglycan binding region
CC and human superoxide dismutase, joined via a linker region was
CC constructed according to the formula M-(Z-M)2-SOD where Z is the peptide
CC -RHHPREMKRVEDL-. The fusion protein is useful for extending
CC the in vivo lifetimes of biologically active cpds. such as SOD and
CC for targetting them to specific cell surfaces or substrates. The
CC glycosaminoglycan (GAG) binding protein is formed into a fusion
CC protein with SOD to increase stability, plasma half-life and ease
CC of purification of SOD. SOD is useful for reduction of tissue damage
CC caused by oxygen radicals and is used in the treatment of autoimmune
CC diseases e.g. rheumatoid and osteo-arthritis.
CC See also R24225-35, R27933-51.
SQ Sequence 205 AA;

Query Match 100.0%; Score 15; DB 1; Length 205;
Best Local Similarity 25.0%; Pred. No. 2.97e+03;
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 6 PREMKRV 13
QY 2 PXXXXXXV 9

RESULT 13
ID R24228 standard; Protein; 212 AA.
AC R24228;
DT 25-NOV-1992 (first entry)
DE GAG fusion protein with SOD according to a formula.
KW Glycosamino:glycan; superoxidismutase; tissue damage;
KW autoimmune disease; rheumatoid arthritis; osteoarthritis; ss.
OS Synthetic.
FH Key Location/Qualifiers
FT peptide 1. .40
FT /note= "GAG binding motif"
FT peptide 41. .212
FT /note= "SOD"
PN WO9207935-A.
PD 14-MAY-1992.
PF 01-NOV-1991: U08105.
PR 01-NOV-1990: US-608539.
PR 02-NOV-1990: US-608569.
PA (SCRI ) SCRIPPS RES INST.
PI Boissinot M, Fisher C, Griffin JH, Hallewell RA, Kuhn L;
PI Mullenbachgt, Parge HE, Tainer JA;
DR WPI: 92-183671/22.
PT Fusion proteins with glycosamino:glycan-binding and
PT superoxidismutase activities - reduce tissue damage caused by
PT super:oxide radicals, useful in treating autoimmune diseases e.g.
PT rheumatoid arthritis and osteoarthritis
PS Claim 7; Fig 1: 140pp; English.
CC The fusion protein comprising the a glycosaminoglycan binding region
CC and human superoxide dismutase, joined via a linker region was
CC constructed according to the formula M-(Z-M)2-SOD where Z is the peptide
CC -RVPSGKKRKRRLKPS-. The fusion protein is useful for extending
CC the in vivo lifetimes of biologically active cpds. such as SOD and

CC for targetting them to specific cell surfaces or substrates. The
CC glycosaminoglycan (GAG) binding protein is formed into a fusion
CC protein with SOD to increase stability, plasma half-life and ease
CC of purification of SOD. SOD is useful for reduction of tissue damage
CC caused by oxygen radicals and is used in the treatment of autoimmune
CC diseases e.g. rheumatoid and osteo-arthritis.
CC See also R24225-35, R27933-51.
SQ Sequence 212 AA;

Query Match 100.0%; Score 15; DB 1; Length 212;
Best Local Similarity 25.0%; Pred. No. 2.97e+03;
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 115 PKDEEHV 122
QY 2 PXXXXXXV 9

RESULT 14
ID R20805 standard; Protein; 220 AA.
AC R20805;
DT 21-MAY-1992 (first entry)
DE Human CD28 antigen.
KW cloning technique; cell surface antigen; immunodiagnosis;
KW tumour.
OS Homo sapiens.
FH Key Location/Qualifiers
FT peptide 1. .18
FT /label= signal
FT protein 19. .220
FT modified_site 37. .39
FT /label= N-linked_glycosylation
FT modified_site 71. .73
FT /label= N-linked_glycosylation
FT modified_site 92. .94
FT /label= N-linked_glycosylation
FT modified_site 105. .107
FT /label= N-linked_glycosylation
FT modified_site 129. .131
FT /label= N-linked_glycosylation
FT region 153. .179
FT /label= transmembrane
PN WO9201049-A.
PD 23-JAN-1992.
PF 15-JUL-1990; U04986.
PR 13-JUL-1990: US-553759.
PA (GHERO-) GEN HOSPITAL CORP.
PI Seed B, Aruffo A, Amlot M;
DR WPI: 92-056864/07.
DR N-PSDB; Q21167.
PT New CD53 cell surface antigen and DNA encoding it - for
PT immuno-therapy and diagnosis of haematopoietic neoplasms, etc.
PS Example 3; Fig 7; 160pp; English.
CC The CD28 antigen amino acid sequence was predicted from the
CC nucleotide sequence of a cDNA clone isolated from a human
CC lymphoblastoid cell line JY library using the antibody enrichment
CC method (see Q21167). The sequence of the CD28 antigen has
CC substantial homology with mouse and rabbit immunoglobulin
CC heavy-chain variable regions over a domain spanning almost the
CC entire extracellular portion of CD28.
SQ Sequence 220 AA;

Query Match 100.0%; Score 15; DB 1; Length 220;
Best Local Similarity 25.0%; Pred. No. 2.97e+03;
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 152 PFWLVVY 159
QY 2 PXXXXXXV 9

RESULT 15
ID R27934 standard; Protein; 221 AA.
```



PT all chains identical and contg. haem, useful as blood substitute  
 PT for transfusion  
 PS Disclosure; Fig 1; 41pp; French.  
 CC Chimeric alpha-beta globin molecules are claimed which, when  
 CC associated together to reconstitute the alpha-beta2 interface and  
 CC incorporating haem, are suitable as blood substitutes. Pref. the  
 CC chimeric chains contain amino acids 1-73 of beta globin at the  
 CC N-terminus and amino acids 69 onwards of alpha-globin at the  
 CC C-terminus. Alternatively, each chain may contain the N-terminus of  
 CC an alpha chain with the C-terminus of a beta-chain, or all 4  
 CC chains can be of the beta type. The tetramers are characterised by  
 CC a lower oxygen affinity than natural haemoglobin.  
 SQ Sequence 141 AA;

Query Match 100.0%; Score 15; DB 1; Length 141;  
 Best Local Similarity 25.0%; Pred. No. 2.97e+03;  
 Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 114 PAFETPAV 121  
 QY 2 PXXXXXXV 9

RESULT 9  
 ID R10474 standard; Protein; 143 AA.  
 AC R10474;  
 DE 10-APR-1991 (first entry)  
 DT Synthetic signal peptide and phleomycin-resistance protein Sh.  
 KW Streptoalloteichus hindustanus; phleomycin; Tolypocladium fungus.  
 PN W09100357-A.  
 PD 10-JAN-1991.  
 PF 28-JUN-1990; F00479.  
 PR 30-JUN-1989; FR-008838.  
 PA (CAYL-) CAYLA.  
 PI Calmels T, Durand H;  
 DR WPI: 91-036751/05.  
 DR N-PSDB; Q10354.  
 PT Recombinant protein prodn. in Tolypocladium fungal strain -  
 PT contg. appropriate DNA pref. including signal sequence, providing  
 PT high yield in simple culture medium  
 PS Claim 19; Fig 2; 41pp; French.  
 CC The Sh protein is the phleomycin-resistance protein from  
 CC Streptoalloteichus hindustanus. Production of this protein in high  
 CC yield is achieved by coupling the sequence coding for the protein to  
 CC a synthetic signal sequence to ensure efficient secretion of the  
 CC protein. The sequences are placed under the control of a fungal  
 CC promoter (see Q10355) in a plasmid containing an ARS and capable of  
 CC replication in a Tolypocladium fungus. The Sh coding sequence also  
 CC acts as its own marker gene when grown on medium containing  
 CC phleomycin. See also Q10356.  
 SQ Sequence 143 AA;

Query Match 100.0%; Score 15; DB 1; Length 143;  
 Best Local Similarity 25.0%; Pred. No. 2.97e+03;  
 Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 28 PVLTAADV 35  
 QY 2 PXXXXXXV 9

RESULT 10  
 ID P82757 standard; protein; 164 AA.  
 AC P82757;  
 DT 11-DEC-1990 (first entry)  
 DE Human colony stimulating factor-1 analogue.  
 KW Colony stimulating factor-1; trans-membrane region.  
 OS Synthetic.  
 FH Key Location/Qualifiers  
 PN EP-249477-A.  
 PD 16-DEC-1987.  
 PF 11-JUN-1987; 305173.  
 PR 12-JUN-1986; US-873620.

PR 12-JUN-1986; US-873497.  
 PA (IMMU-) IMMUNEX CORP.  
 PI Cerretti DP, Clevenger WR, Cosman DU, Gimpel SD, Price VL,  
 PI Urdal DL;  
 DR WPI: 88-022635/04.  
 DR N-PSDB; N82206.  
 PT Recombinant protein analogues - with modified trans-membrane  
 PT region to facilitate expression by host cells  
 PS Disclosure; P; English.  
 CC This protein analogue has a sequence which differs from that of  
 CC natural human colony stimulating factor (CSF)-1 at several posns.  
 CC The modifications constitute deletions or substitutions of hydro-  
 CC phobic amino acids in the transmembrane domain. The protein is  
 CC expressed by host cells. This allows biologically active CSF to be  
 CC produced in commercially viable amts. by large scale yeast ferment-  
 CC ation.  
 SQ Sequence 164 AA;

Query Match 100.0%; Score 15; DB 1; Length 164;  
 Best Local Similarity 25.0%; Pred. No. 2.97e+03;  
 Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 2 PLDKREEV 9  
 QY 2 PXXXXXXV 9

RESULT 11  
 ID R13514 standard; Protein; 174 AA.  
 AC R13514;  
 DT 25-OCT-1991 (first entry)  
 DE P.denitrificans COB P.  
 KW cob gene; corrinoid; descobalocorrinoid; cor gene.  
 OS Pseudomonas denitrificans.  
 PN W09111518-A.  
 PD 08-AUG-1991.  
 PF 30-JAN-1991; F00054.  
 PR 31-JAN-1990; FR-001137.  
 PA (RHON) RHONE-POULENC BIOCH.  
 PI Blanche F, Meron B, Crouzet J, Debussche L, Levy-Schil S;  
 PI Thibaut D;  
 DR WPI: 91-252650/34.  
 DR N-PSDB; Q13288.  
 PT New polypeptide(s) involved in cobalamin and cobamide  
 PT biosynthesis - and DNA encoding them, for amplification of  
 PT cobalamin, esp. coenzyme B12 prodn.  
 PS Claim 39; Fig 47; 299pp; French.  
 CC This sequence corresponds to one of 24 polypeptides obtained from  
 CC P.denitrificans and implicated in the biosynthesis of cobalamines  
 CC and/or cobamides. It has cobinamide kinase and cobinamide phosphate  
 CC guanylttransferase activity and is encoded by part of a 13144 bp  
 CC fragment isolated from a P.denitrificans genomic DNA bank constructed  
 CC in vector pXL59. See also Q13284-Q13287.  
 SQ Sequence 174 AA;

Query Match 100.0%; Score 15; DB 1; Length 174;  
 Best Local Similarity 25.0%; Pred. No. 2.97e+03;  
 Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 116 PEARARLV 123  
 QY 2 PXXXXXXV 9

RESULT 12  
 ID R27932 standard; Protein; 205 AA.  
 AC R27932;  
 DT 25-NOV-1992 (first entry)  
 DE GAG fusion protein with SOB according to a formula.  
 KW Glycosamino-glycan; superoxidisedismutase; tissue damage;  
 KW autoimmune disease; rheumatoid arthritis; osteoarthritis; ss.  
 OS Synthetic.  
 FH Key Location/Qualifiers

CC be used for purifying, detecting and isolating antibodies.  
SQ Sequence 109 AA;

Query Match 100.0%; Score 15; DB 1; Length 109;  
Best Local Similarity 25.0%; Pred. No. 2.97e+03;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 82 PEASIPLV 89  
QY 2 PXXXXXXV 9

## RESULT 6

ID R33679 standard; Protein; 110 AA.  
AC R33679;  
DT 03-JUL-1993 (first entry)  
DE Variant IgE - mutant Emu76.  
KW High affinity; FCEH; low affinity; FCEL; Padian;  
KW IgE receptor; Fc; IgG1.  
OS Homo sapiens.

FH Key Location/Qualifiers

FT region 7..12

FT /label= beta-strand\_A

FT region 13..24

FT /label= loop\_AB

FT region 25..33

FT /label= beta-strand\_B

FT region 34..42

FT /label= loop\_BC

FT region 43..48

FT /label= beta-strand\_C

FT region 49..57

FT /label= loop\_CD

FT region 58..65

FT /label= beta-strand\_D

FT region 66..67

FT /label= loop\_DE

FT region 68..78

FT /label= beta-strand\_E

FT region 79..86

FT /label= loop\_EF

FT region 87..94

FT /label= beta-strand\_F

FT region 95..100

FT /label= loop\_FG

FT region 101..105

FT /label= beta-strand\_G

FT misc\_difference 62..64

FT /label= mutation

FT /note= "KQR -> AAA"

PN WO9304173-A.

PD 04-MAR-1993.

PF 14-AUG-1992; U06860.

PR 14-AUG-1991; US-744768.

PR 07-MAY-1992; US-879495.

PA (GETH ) GENENTECH INC.

PI Jardieu PM, Presta LG;

DR WPI; 93-094004/11.

PT Polypeptide(s) binding to specific Fc epsilon receptors - act as

PT IgE antagonists; useful for treating and preventing IgE-mediated

PT disorders e.g. allergies

PS Disclosure; Page 73; 113pp; English.

CC IgE mutants were prepd. to evaluate their effect on binding to

CC anti-IgE, esp. MAE11, and to Fc epsilon RI and Fc epsilon RII.

CC Some of the mutants were designed to substitute for a specific

CC amino acid residue another residue with either similar or very

CC different charge or size.

SQ Sequence 110 AA;

Query Match 100.0%; Score 15; DB 1; Length 110;  
Best Local Similarity 25.0%; Pred. No. 2.97e+03;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 25 PTITCLV 32  
QY 2 PXXXXXXV 9

## RESULT 7

ID R23770 standard; Protein; 136 AA.

AC R23770;

DT 27-OCT-1992 (first entry)

DE Recombinant light chain variable domain (4).

KW Complementarity determining region; light chain variable domain;

KW antigen binding site; ligand; framework region; cancer; transplant.

OS Synthetic.

FH Key Location/Qualifiers

FT region 23..35

FT /label= CDR(d)

FT region 51..63

FT /label= CDR(d)

FT region 98..106

FT /label= CDR(f)

PN WO9206193-A.

PD 16-APR-1992.

PF 04-OCT-1991; G01726.

PR 05-OCT-1990; GB-021679.

PA (GORMA/) GORMAN S D.

PI Gorman SD, Routledge EG, Waldmann H;

DR WPI; 92-150879/18.

PT Ligands and antibodies with binding affinity for CD3 antigen -

PT for treatment of immunosuppression e.g. in graft rejection, and

PT cancer, esp. lymphoid malignancies

PS Claim 7; Page 31; 49pp; English.

CC The sequence given is a recombinant human light chain variable

CC domain ligand containing the complementarity determining region

CC (CDR) given in R23736 and R23738. CDR's are found in the variable

CC and act as connectors between the four framework regions.

CC It has been noted that there seem to be no characteristic features

CC which distinguish human from mouse or rat CDR's and they are

CC therefore immunologically identical. This ligand has binding affinity

CC for the human CD3 antigen and due to the lack of immunological

CC response caused by the synthetic CDR's the ligand can be considered to

CC be humanised. This ligand can be used to manufacture medicaments

CC for use in immunosuppression esp. in patients with cancer or transplant

CC recipients.

SQ Sequence 136 AA;

Query Match 100.0%; Score 15; DB 1; Length 136;

Best Local Similarity 25.0%; Pred. No. 2.97e+03;

Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 41 PGRAPTV 48

QY 2 PXXXXXXV 9

## RESULT 8

ID R42631 standard; Protein; 141 AA.

AC R42631;

DT 26-APR-1994 (first entry)

DE Natural alpha-globin.

KW Haemoglobin; alpha globin; beta globin; blood substitute;

KW oxygen carrier; reversible oxygen fixation; chimeric gene;

KW fusion protein.

PN WO9319089-A.

PD 30-SEP-1993.

PR 18-MAR-1993; F00273.

PR 18-MAR-1992; FR-003224.

PA (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.

PI Edelstein S, Pagnier RJ, Poyart C;

DR WPI; 93-320682/40.

DR N-PSDB; Q49615.

PT New synthetic tetrameric globin type oxygen transporter - with

RESULT 2  
ID R23272 standard; Protein; 84 AA.  
AC R23272;  
DT 22-JUN-1992 (first entry)  
DE Bovine parathyroid hormone analogue, [Ser1Orn3]bPTH.  
KW PTH; agonist; antagonist; receptor binding.  
OS Synthetic.  
FH Key Location/Qualifiers  
FT modified\_site 3  
FT /label= Orn  
PD WO9200753-A.  
PN 23-JAN-1992.  
PR 15-JUL-1990; U04971.  
PR 13-JUL-1990; US-553760.  
PA (REGC ) Univ of California.  
PI Cohen FA, Nissenson RA, Strewler GJ;  
DR WPI; 92-056643/07.  
PT New parathyroid hormone analogues - useful in treating cancer,  
PT osteoporosis, hypercalcaemia and hyper-parathyroid conditions.  
PS Claim 15; Page 63; 86pp; English.  
CC Residues 35-84 may be absent. The C-terminal gp. may be -COOH,  
CC -COO-M (M+ = cation), or -(C-O)NH2. 142 specific peptides derived  
CC from bovine PTH are given in the specification (R23251-382, and  
CC R23540-549). Corresp. peptides created using residues 7-84  
CC of human and porcine PTH are also claimed. All have mutations at  
CC positions 3, and/or 6, and/or 9 which result in surface side chains  
CC which are useful to modulate receptor binding and activity. They  
CC are useful as agonists and antagonists in the treatment of condi-  
CC tions or diseases involving PTH. The peptides are pref. prepd. by  
CC solid phase synthesis. See also R21257 (human generic), R21258  
CC (bovine generic) and R21259 (porcine generic).  
SQ Sequence 84 AA;  
Query Match 100.0%; Score 15; DB 1; Length 84;  
Best Local Similarity 25.0%; Pred. No. 2.97e+03;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
Db 51 PRKKNV 58  
QY 2 PXXXXXXV 9  
RESULT 3  
ID R21158 standard; Protein; 84 AA.  
AC R21158;  
DT 17-JUN-1992 (first entry)  
DE Human parathyroid hormone analogue, [Lys6]hPTH.  
KW PTH; agonist; antagonist; receptor binding.  
OS Synthetic.  
PN WO9200753-A.  
PD 23-JAN-1992.  
PR 13-JUL-1990; US-553760.  
PR 15-JUL-1990; U04971.  
PA (REGC ) Univ of California.  
PI Cohen FA, Nissenson RA, Strewler GJ;  
DR WPI; 92-056643/07.  
PT New parathyroid hormone analogues - useful in treating cancer,  
PT osteoporosis, hypercalcaemia and hyper-parathyroid conditions.  
PS Claim 15; Page 63; 86pp; English.  
CC Residues 35-84 may be absent. The C-terminal gp. may be -COOH,  
CC -COO-M (M+ = cation), or -(C-O)NH2. 140 specific peptides derived  
CC from human PTH are given in the specification (R21150-256, R23226-  
CC 3250 + R23522-529). Corresp. peptides created using residues 7-84  
CC of bovine and porcine PTH are also claimed. All have mutations at  
CC positions 3, and/or 6, and/or 9 which result in surface side chains  
CC which are useful to modulate receptor binding and activity. They  
CC are useful as agonists and antagonists in the treatment of condi-  
CC tions or diseases involving PTH. The peptides are pref. prepd. by  
CC solid phase synthesis. See also R21257 (human generic), R21258  
CC (bovine generic) and R21259 (porcine generic).  
SQ Sequence 84 AA;  
Query Match 100.0%; Score 15; DB 1; Length 84;  
Best Local Similarity 25.0%; Pred. No. 2.97e+03;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
Db 51 PRKKNV 58  
QY 2 PXXXXXXV 9  
RESULT 3  
ID R21158 standard; Protein; 84 AA.  
AC R21158;  
DT 17-JUN-1992 (first entry)  
DE Human parathyroid hormone analogue, [Lys6]hPTH.  
KW PTH; agonist; antagonist; receptor binding.  
OS Synthetic.  
PN WO9200753-A.  
PD 23-JAN-1992.  
PR 13-JUL-1990; US-553760.  
PR 15-JUL-1990; U04971.  
PA (REGC ) Univ of California.  
PI Cohen FA, Nissenson RA, Strewler GJ;  
DR WPI; 92-056643/07.  
PT New parathyroid hormone analogues - useful in treating cancer,  
PT osteoporosis, hypercalcaemia and hyper-parathyroid conditions.  
PS Claim 15; Page 63; 86pp; English.  
CC Residues 35-84 may be absent. The C-terminal gp. may be -COOH,  
CC -COO-M (M+ = cation), or -(C-O)NH2. 140 specific peptides derived  
CC from human PTH are given in the specification (R21150-256, R23226-  
CC 3250 + R23522-529). Corresp. peptides created using residues 7-84  
CC of bovine and porcine PTH are also claimed. All have mutations at  
CC positions 3, and/or 6, and/or 9 which result in surface side chains  
CC which are useful to modulate receptor binding and activity. They  
CC are useful as agonists and antagonists in the treatment of condi-  
CC tions or diseases involving PTH. The peptides are pref. prepd. by  
CC solid phase synthesis. See also R21257 (human generic), R21258  
CC (bovine generic) and R21259 (porcine generic).  
SQ Sequence 84 AA;  
Query Match 100.0%; Score 15; DB 1; Length 84;  
Best Local Similarity 25.0%; Pred. No. 2.97e+03;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
Db 51 PRKKNV 58  
QY 2 PXXXXXXV 9

Best Local Similarity 25.0%; Pred. No. 2.97e+03;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
Db 51 PRKKNV 58  
QY 2 PXXXXXXV 9  
RESULT 4  
ID R21159 standard; Protein; 84 AA.  
AC R21159;  
DT 17-JUN-1992 (first entry)  
DE Human parathyroid hormone analogue, [Arg6]hPTH.  
KW PTH; agonist; antagonist; receptor binding.  
OS Synthetic.  
PN WO9200753-A.  
PD 23-JAN-1992.  
PR 15-JUL-1990; U04971.  
PR 13-JUL-1990; US-553760.  
PA (REGC ) Univ of California.  
PI Cohen FA, Nissenson RA, Strewler GJ;  
DR WPI; 92-056643/07.  
PT New parathyroid hormone analogues - useful in treating cancer,  
PT osteoporosis, hypercalcaemia and hyper-parathyroid conditions.  
PS Claim 15; Page 63; 86pp; English.  
CC Residues 35-84 may be absent. The C-terminal gp. may be -COOH,  
CC -COO-M (M+ = cation), or -(C-O)NH2. 140 specific peptides derived  
CC from human PTH are given in the specification (R21150-256, R23226-  
CC 3250 + R23522-529). Corresp. peptides created using residues 7-84  
CC of bovine and porcine PTH are also claimed. All have mutations at  
CC positions 3, and/or 6, and/or 9 which result in surface side chains  
CC which are useful to modulate receptor binding and activity. They  
CC are useful as agonists and antagonists in the treatment of condi-  
CC tions or diseases involving PTH. The peptides are pref. prepd. by  
CC solid phase synthesis. See also R21257 (human generic), R21258  
CC (bovine generic) and R21259 (porcine generic).  
SQ Sequence 84 AA;  
Query Match 100.0%; Score 15; DB 1; Length 84;  
Best Local Similarity 25.0%; Pred. No. 2.97e+03;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
Db 51 PRKKNV 58  
QY 2 PXXXXXXV 9  
RESULT 5  
ID R07009 standard; Protein; 109 AA.  
AC R07009;  
DT 17-JAN-1991 (first entry)  
DE Protein G variant.  
KW Immunoglobulin.  
OS Streptococcus sp. Lancefield Group G strain.  
PN US4956296-A.  
PD 11-SEP-1990.  
PR 20-JUN-1988; 209236.  
PR 14-FEB-1986; US-829354.  
PR 23-APR-1986; US-854887.  
PR 17-FEB-1987; WO-U00329.  
PR 19-JUN-1987; US-063959.  
PR 20-JUN-1988; US-209236.  
PA (GENE-) GENEX CORP.  
PI Farnestock SR;  
DR WPI; 90-297491/39.  
DR N-PSDB; Q06014.  
PT Recombinant Protein G variants - obt'd. using a cloned gene  
PT encoding Protein G from Streptococcus sp., used for binding  
PT immunoglobulin.  
PS Disclosure; Column 9-16; 48pp; English.  
CC Sequence may be incorporated into a non-pathogenic host eg. E.coli,  
CC where they may be expressed at high levels. The proteins have a  
CC higher binding efficiency and capacity for immunoglobulin, and may

\*\*\*\*\*

W P S R E H (TM)

\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MParch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Sat Apr 15 01:55:10 2000; MasPar time 3.10 Seconds  
Tabular output not generated. 68.745 Million cell updates/sec

Title: >US-08-452-843-27  
Description: (1-9) from US08452843.pep  
Perfect Score: 15  
Sequence: 1 PXXXXXXV 9

Scoring table: PAM 150  
Gap 15

Searched: 188963 seqs, 23686106 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: a:geneseq36  
1:geneseq9

Statistics: Mean 7.455; Variance 9.377; scale 0.795

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description	Pred. No.
1	15	100.0	56	R12114	17kD TNF linked to gam	2.97e+03
2	15	100.0	84	R23272	Bovine parathyroid hor	2.97e+03
3	15	100.0	84	R21158	Human parathyroid horm	2.97e+03
4	15	100.0	84	R21159	Human parathyroid horm	2.97e+03
5	15	100.0	109	R07009	Protein G variant.	2.97e+03
6	15	100.0	110	R33679	Variant Ige - mutant E	2.97e+03
7	15	100.0	136	R23770	Recombinant light chai	2.97e+03
8	15	100.0	141	R42631	Natural alpha-globin.	2.97e+03
9	15	100.0	143	R10474	Synthetic signal pepti	2.97e+03
10	15	100.0	164	R82757	Human colony stimulati	2.97e+03
11	15	100.0	174	R13514	P.denitrificans COB P.	2.97e+03
12	15	100.0	205	R27932	GAG fusion protein wit	2.97e+03
13	15	100.0	212	R24228	GAG fusion protein wit	2.97e+03
14	15	100.0	220	R20805	Human CD28 antigen.	2.97e+03
15	15	100.0	221	R27934	GAG fusion protein wit	2.97e+03
16	15	100.0	228	R24227	GAG fusion protein wit	2.97e+03
17	15	100.0	231	R12606	TNF mutein for virion	2.97e+03
18	15	100.0	263	P90507	Sequence of an epitope	2.97e+03
19	15	100.0	269	R10445	S014 mutant of the sub	2.97e+03
20	15	100.0	274	R81275	Human alpha 2-plasmin	2.97e+03
21	15	100.0	291	R22638	Generic sequence of Hu	2.97e+03
22	15	100.0	299	R05369	Protein coded for by t	2.97e+03
23	15	100.0	300	R05109	Sequence encoded by RN	2.97e+03

24	15	100.0	321	1	R22045	Carcino embryonic anti	2.97e+03
25	15	100.0	329	1	R05371	IRF-1 active protein.	2.97e+03
26	15	100.0	330	1	R29926	Eimeria antigen Eam45	2.97e+03
27	15	100.0	345	1	P83149	Probe F10-encoded prot	2.97e+03
28	15	100.0	361	1	P90506	Sequence of an epitope	2.97e+03
29	15	100.0	405	1	R28840	HeLa cell fucosyltrans	2.97e+03
30	15	100.0	416	1	R24526	HCV in expression vect	2.97e+03
31	15	100.0	431	1	P71663	Modified prourokinase.	2.97e+03
32	15	100.0	436	1	R22513	Truncated precursor of	2.97e+03
33	15	100.0	455	1	R20787	TNF-alpha binding prot	2.97e+03
34	15	100.0	464	1	R10389	Antithrombin III mutan	2.97e+03
35	15	100.0	464	1	R42921	Human antithrombin III	2.97e+03
36	15	100.0	464	1	R42926	Human antithrombin III	2.97e+03
37	15	100.0	467	1	R41890	Bile acid sulphate sul	2.97e+03
38	15	100.0	488	1	R22512	Mutated precursor of h	2.97e+03
39	15	100.0	494	1	R24400	Recombinant thrombin-b	2.97e+03
40	15	100.0	508	1	P70666	Sequence encoded by LA	2.97e+03
41	15	100.0	581	1	R13490	Human C4 binding prote	2.97e+03
42	15	100.0	768	1	R29850	HSA.	2.97e+03
43	15	100.0	864	1	R24042	HCV NS2-NS4 peptide N2	2.97e+03
44	15	100.0	864	1	R24042	Lipoxygenase.	2.97e+03
45	15	100.0	1967	1	R33547	Sequence of the alpha	2.97e+03

ALIGNMENTS

RESULT 1  
ID R12114 standard; Protein; 56 AA.  
AC R12114;  
DT 30-JUL-1991 (first entry)  
DE 17kD TNF linked to gamma-IFN signal peptide.  
KW Gamma interferon; IFN; tumour necrosis factor; TNF; virion;  
KW drug delivery; tumour; leukaemia.  
FH Key Location/Qualifiers  
FT Peptide I.:20  
FT protein /label= gamma-IFN-sig\_peptide  
FT /label= 21..56  
FT /label= TNF-N-terminal  
PN W09106658-A.  
PD 16-MAY-1991.  
PF 23-OCT-1990; U06141.  
PR 24-OCT-1989; US-426986.  
PR 02-FEB-1990; US-474169.  
PR 02-MAR-1990; US-488706.  
PA (CETU ) CETUS CORP.  
PI Krieglner M. Perez CF;  
DR WPI; 91-164207/22.  
DR N-PSDB; Q11831.  
PT Virion comprising an RNA genome useful for drug delivery system -  
PT contains e.g. IL-2 multiple drug resistance or TNF encoding  
sequences, for treating solid tumours, leukaemia etc.  
PS Disclosure; Fig 1: 59pp; English.  
CC The TNF-encoding fragment is inserted into pFVX (contg. LTR's  
CC and psi site of MoMuV where the splice donor site is replaced by  
CC an HsV sequence). pFVX-TNF can be used to transfect suitable  
CC packaging cells or its 26kD TNF-encoding fragment is cut out and  
CC mutagenised with an oligonucleotide encoding the gamma-IFN  
CC signal peptide. The sequence is introduced into  
CC pUC.FVXdeltaIIIIRNgammaasig. NIH-3T3 cells transfected with  
CC this construct revealed no detectable 26kD, but considerable  
CC amounts of the 17kD TNF in the lysate and supernatant.  
CC See also Q11832.  
SQ Sequence 56 AA;

Query Match 100.0%; Score 15; DB 1; Length 56;  
Best Local Similarity 25.0%; Pred. No. 2.97e+03;  
Matches 2; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
Db 28 PEGTFFPV 35  
QY 2 PXXXXXXV 9

Best Local Similarity 36.4%; Pred. No. 9.19e+01;  
Matches 4; Conservative 0; Mismatches 7; Indels 0; Gaps 0;  
Db 374 PTNLKSLNFWY 384  
|  
|  
|  
Oy 2 PXXXXXXFWY 12

Search completed: Sat Apr 15 01:44:54 2000  
Job time : 93 secs.

RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;  
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
RL elegans";  
RL Nature 368:32-38(1994).  
DR EMBL: Z80216; CAB02287.1; -.  
DR PROSITE: PS00383; TYR\_PHOSPHATASE\_1; 1.  
KW Hydrolase.  
SQ SEQUENCE 352 AA; 40502 MW; 02C37A7F CRC32;

Query Match 100.0%; Score 49; DB 5; Length 352;  
Best Local Similarity 36.4%; Pred. No. 9.19e+01;  
Matches 4; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Db 128 PLEKTCADFWY 138  
|  
|  
|  
QY 2 PXXXXXXFWY 12

RESULT 12  
ID Q9XJ40 PRELIMINARY; PRT; 356 AA.  
AC Q9XJ40;  
DT 01-NOV-1999 (TrEMBLrel. 12, Created)  
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)  
DE CHLOROPHYLL B SYNTHASE (FRAGMENT).  
GN CAO.  
OS Oryza sativa (Rice).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;  
OC Poaceae; Oryza.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 99334926.  
RA TOMITANI A., OKADA K., MIYASHITA H., MATTHIJS H.C.P., OHNO T.,  
RA TANAKA A.;  
RT "Chlorophyll b and phycobillins in the common ancestor of cyanobacteria  
RT and chloroplasts";  
RL Nature 400:159-162(1999).  
DR EMBL: AB021310; BAA82479.1; -.  
FT NON\_TER 1  
SQ SEQUENCE 356 AA; 40532 MW; 1D302129 CRC32;

Query Match 100.0%; Score 49; DB 10; Length 356;  
Best Local Similarity 36.4%; Pred. No. 9.19e+01;  
Matches 4; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Db 26 PYNPSLKNFWY 36  
|  
|  
|  
QY 2 PXXXXXXFWY 12

RESULT 13  
ID O70169 PRELIMINARY; PRT; 367 AA.  
AC O70169;  
DT 01-AUG-1998 (TrEMBLrel. 07, Created)  
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)  
DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)  
DE TESTICULAR SERINE PROTEASE 1 (TESP1).  
GN TESP1.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 98249770.  
RA KOHNO N., YAMAGATA K., YAMADA S., KASHIWABARA S., SAKAI Y., BABA T.;  
RT "Two novel testicular serine proteases, TESP1 and TESP2, are present  
RT in the mouse sperm acrosome";  
RL Biochem. Biophys. Res. Commun. 245:658-665(1998).  
DR EMBL: AB008910; BAA26132.1; -.  
DR HSSP: P00766; ICHG.  
DR MGD: MGI:1270856; Tespl.  
DR PFAM: PF00089; trypsin; 1.

DR PRINTS; PR00722; CHYMOTRYPSIN.  
SQ SEQUENCE 367 AA; 40765 MW; 11E86AEE CRC32;

Query Match 100.0%; Score 49; DB 11; Length 367;  
Best Local Similarity 36.4%; Pred. No. 9.19e+01;  
Matches 4; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Db 266 PLVCFLNSEFWY 276  
|  
|  
|  
QY 2 PXXXXXXFWY 12

RESULT 14  
ID P74455 PRELIMINARY; PRT; 372 AA.  
AC P74455;  
DT 01-FEB-1997 (TrEMBLrel. 02, Created)  
DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)  
DT 01-FEB-1999 (TrEMBLrel. 09, Last annotation update)  
DE HIGH-AFFINITY BRANCHED-CHAIN AMINO ACID TRANSPORT PROTEIN BRAB.  
OS Synechocystis sp. (strain PCC 6803).  
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-PCC6803;  
RA TABATA S.;  
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-PCC6803;  
RX MEDLINE: 97061201.  
RA KANEKO T., SATO S., KOTANI H., TANAKA A., ASAMIZU E., NAKAMURA Y.,  
RA MIYAJIMA N., HIROSAWA M., SUGIURA M., SASAMOTO S., KIMURA T.,  
RA HOSOUCHI T., MATSUNO A., MURAKI A., NAKAZAKI N., NARUO K., OKUMURA S.,  
RA SHIMPO S., TAKEUCHI C., WADA T., WATANABE A., YAMADA M., YASUDA M.,  
RA TABATA S.;  
RT "Sequence analysis of the genome of the unicellular cyanobacterium  
RT Synechocystis sp. strain PCC6803. II. Sequence determination of the  
RT entire genome and assignment of potential protein-coding regions.";  
RL DNA Res. 3:109-136(1996).  
DR EMBL: D90915; BAA18556.1; -.  
SQ SEQUENCE 372 AA; 40826 MW; FBC015C7 CRC32;

Query Match 100.0%; Score 49; DB 2; Length 372;  
Best Local Similarity 36.4%; Pred. No. 9.19e+01;  
Matches 4; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Db 245 PRALGKNVFWY 255  
|  
|  
|  
QY 2 PXXXXXXFWY 12

RESULT 15  
ID O44371 PRELIMINARY; PRT; 407 AA.  
AC O44371;  
DT 01-JUN-1998 (TrEMBLrel. 06, Created)  
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)  
DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)  
DE UNC-129.  
GN UNC-129.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;  
OC Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-N2;  
RA COLAVITA A., KRISNA S., ZHENG H., PADGETT R.W., CULOTTI J.G.;  
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AF029887; AAC48376.1; -.  
DR HSSP: P18075; 1BMP.  
DR PFAM: PF00019; TGF-beta; 1.  
SQ SEQUENCE 407 AA; 47338 MW; C8F6C1CE CRC32;

Query Match 100.0%; Score 49; DB 5; Length 407;

RX MEDLINE; 92167809.  
RA VENKATESAN M.M., BUSSE J.M., HARTMAN A.B.;  
RT "Sequence variation in two ipaH genes of Shigella flexneri 5 and  
RT homology to the LRG-like family of proteins."  
RL Mol. Microbiol. 5:2435-2445(1991).  
[2]  
RN SEQUENCE FROM N.A.  
RP STRAIN-M9OT-W, SEROTYPE 5;  
RX MEDLINE: 97074644.  
RA VENKATESAN M.M., ALEXANDER W.A., FERNANDEZ-PRADA C.;  
RT "A Shigella flexneri invasion plasmid gene, ipaH, with homology to  
RT IS29 and sequences encoding bacterial sugar phosphate transport  
RT proteins."  
RL Gene 175:23-27(1996).  
DR EMBL; U28354; AAC44575.1; -.  
KW Plasmid.  
SQ SEQUENCE 333 AA; 36475 MW; 0EB41D70 CRC32;

Query Match 100.0%; Score 49; DB 2; Length 333;  
Best Local Similarity 36.4%; Pred. No. 9.19e+01;  
Matches 4; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Db 39 PPGAKTLVFWY 49  
QY 2 PXXXXXXFWY 12

RESULT 9  
ID Q9YFX4 PRELIMINARY; PRT; 339 AA.  
AC Q9YFX4;  
DT 01-NOV-1999 (TRENBLrel. 12, Created)  
DT 01-NOV-1999 (TRENBLrel. 12, Last sequence update)  
DT 01-NOV-1999 (TRENBLrel. 12, Last annotation update)  
DE 339AA LONG HYPOTHETICAL PROTEIN.  
GN AP00126.  
OS Aeropyrum pernix.  
OC Archaea; Crenarchaeota; Aeropyrum.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-K1;  
RX MEDLINE: 99310339.  
RA KAWABAYASHI Y., HINO Y., HORIKAWA H., YAMAZAKI S., HAIKAWA Y.,  
RA JIN-NO K., TAKAHASHI M., SEKINE M., BABA S., ANKAI A., KOSUGI H.,  
RA HOSOIYAMA A., FUKUI S., NAGAI Y., NISHIJIMA K., NAKAZAWA H.,  
RA TAKAMIYA M., MASUDA S., FUNAHASHI T., TANAKA T., KUDOH Y.,  
RA YAMAZAKI J., KUSHIDA N., OGUCHI A., AOKI K., KUBOTA K., NAKAMURA Y.,  
RA NOMURA N., SAKO Y., KIKUCHI H.;  
RT "Complete genome sequence of an aerobic hyper-thermophilic  
RT Crenarchaeon, Aeropyrum pernix K1."  
RL DNA Res. 6:83-101(1999).  
DR EMBL; AP000058; BAA79037.1; -.  
SQ SEQUENCE 339 AA; 36227 MW; 5AD6A7C2 CRC32;

Query Match 100.0%; Score 49; DB 1; Length 339;  
Best Local Similarity 36.4%; Pred. No. 9.19e+01;  
Matches 4; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Db 201 PSLGEVRGFWY 211  
QY 2 PXXXXXXFWY 12

RESULT 10  
ID O44901 PRELIMINARY; PRT; 344 AA.  
AC O44901;  
DT 01-JUN-1998 (TRENBLrel. 06, Created)  
DT 01-JUN-1998 (TRENBLrel. 06, Last sequence update)  
DT 01-NOV-1999 (TRENBLrel. 12, Last annotation update)  
DE 2K484.7 PROTEIN (EC 3.1.3.48).  
GN 2K484.7.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;  
OC Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.

RN SEQUENCE FROM N.A.  
RP STRAIN-BRISTOL N2;  
RX MEDLINE: 94150718.  
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,  
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,  
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,  
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,  
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,  
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,  
RA PARSONS J., PERCY C., RIPPEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,  
RA SWALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,  
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,  
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;  
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
RT elegans."  
RL Nature 368:32-38(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RX MURRAY J., WOHLDMANN P.;  
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.  
[3]  
RN SEQUENCE FROM N.A.  
RP STRAIN-BRISTOL N2;  
RX WATERSTON R.;  
RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF040659; AAB95064.1; -.  
DR HSP; P18052; IYFO  
DR PROSITE; PS00383; TYR\_PHOSPHATASE\_1; 1.  
DR PFAM; PF00102; Y\_phosphatase; 1.  
KW Hydrolase.  
SQ SEQUENCE 344 AA; 39516 MW; D71CBE01 CRC32;

Query Match 100.0%; Score 49; DB 5; Length 344;  
Best Local Similarity 36.4%; Pred. No. 9.19e+01;  
Matches 4; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Db 127 PLDKSCADFWY 137  
QY 2 PXXXXXXFWY 12

RESULT 11  
ID Q9XVU4 PRELIMINARY; PRT; 352 AA.  
AC Q9XVU4;  
DT 01-NOV-1999 (TRENBLrel. 12, Created)  
DT 01-NOV-1999 (TRENBLrel. 12, Last sequence update)  
DT 01-NOV-1999 (TRENBLrel. 12, Last annotation update)  
DE F10G8.1 PROTEIN (EC 3.1.3.48).  
GN F10G8.1.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;  
OC Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX BASHAM V.;  
RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.  
[2]  
RN SEQUENCE FROM N.A.  
RX MEDLINE: 94150718.  
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,  
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,  
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,  
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,  
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,  
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,  
RA PARSONS J., PERCY C., RIPPEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,  
RA SWALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,  
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,

RA MATSUO T., OSMI-YAMASHITA N., NOJI S., OHUCHI H., KOYAMA E.,  
RA MYOKAI F., MATSUO N., TANIGUCHI S., DOI H., ISEKI S., NINOMIYA Y.,  
RA FUJIWARA M., WANTANABE T., ETO K.;  
RT "A mutation in the Pax-6 gene in rat small eye is associated with  
RT impaired migration of midbrain crest cells.";  
RL Nat. Genet. 3:299-304(1993).  
CC -|- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).  
DR EMBL: S74393; AAB32671.1; -;  
DR HSP: P06601; 1FJL.  
DR PROSITE: PS00027; HOMEBOX\_1; 1.  
DR PFAM: PF00046; homeobox; 1.  
DR PFAM: PF00292; PAX; 1.  
KW Homeobox; DNA-binding; Nuclear protein.  
FT NON\_TER 1  
SQ SEQUENCE 269 AA; 31098 MW; 44PFC4C CRC32;  
Query Match 100.0%; Score 49; DB 11; Length 269;  
Best Local Similarity 36.4%; Pred. No. 9.19e+01;  
Matches 4; Conservative 0; Mismatches 7; Indels 0; Gaps 0;  
Db 193 PISWNKSEFWY 203  
QY 2 PXXXXXXFWY 12  
RESULT 6  
ID Q18415 PRELIMINARY; PRT; 288 AA.  
AC Q18415;  
DT 01-NOV-1996 (TRENBLrel. 01, Created)  
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)  
DT 01-NOV-1999 (TRENBLrel. 12, Last annotation update)  
DE SIMILAR TO PROTEIN-TYROSINE PHOSPHATASE.  
DE NCBI GI: 1086598 (EC 3.1.3.48).  
GN C3H5.16.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;  
OC Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RX MEDLINE: 94150718.  
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,  
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,  
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,  
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,  
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,  
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,  
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,  
RA SMAILDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,  
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,  
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;  
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
RT elegans.";  
RL Nature 368:32-38(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RX MEDLINE: 94150718.  
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,  
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,  
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,  
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,  
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,  
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,  
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,  
RA SMAILDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,  
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,  
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;  
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
RT elegans.";  
RL Nature 368:32-38(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RA BRADSHAW H., STELLIES L.;  
RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RA WATERSTON R.;  
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.  
DR EMBL: U41007; AAB82276.1; -;  
DR HSP: P18052; IYFO.  
DR PROSITE: PS00383; TYR\_PHOSPHATASE\_1; 1.  
DR PFAM: PF00102; Y\_phosphatase; 1.  
KW Hydrolase.  
SQ SEQUENCE 288 AA; 33608 MW; AB7EC854 CRC32;  
Query Match 100.0%; Score 49; DB 5; Length 288;  
Best Local Similarity 36.4%; Pred. No. 9.19e+01;  
Matches 4; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Matches 4; Conservative 0; Mismatches 7; Indels 0; Gaps 0;  
Db 108 PLDNTCADEFWY 118  
QY 2 PXXXXXXFWY 12  
RESULT 7  
ID O16235 PRELIMINARY; PRT; 317 AA.  
AC O16235;  
DT 01-JAN-1998 (TRENBLrel. 05, Created)  
DT 01-JAN-1998 (TRENBLrel. 05, Last sequence update)  
DT 01-NOV-1999 (TRENBLrel. 12, Last annotation update)  
DE F07G11.1 PROTEIN.  
GN F07G11.1.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;  
OC Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RX MEDLINE: 94150718.  
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,  
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,  
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,  
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,  
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,  
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,  
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,  
RA SMAILDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,  
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,  
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;  
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
RT elegans.";  
RL Nature 368:32-38(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RA SAMMONS L., WOHLDMANN P., SANSONE J.;  
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RA WATERSTON R.;  
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AF016419; AAB65297.1; -;  
DR PFAM: PF01482; DUF13.1;  
SQ SEQUENCE 317 AA; 37341 MW; 80F8A64D CRC32;  
Query Match 100.0%; Score 49; DB 5; Length 317;  
Best Local Similarity 36.4%; Pred. No. 9.19e+01;  
Matches 4; Conservative 0; Mismatches 7; Indels 0; Gaps 0;  
Db 74 PLKGQHKFEFWY 84  
QY 2 PXXXXXXFWY 12  
RESULT 8  
ID Q54147 PRELIMINARY; PRT; 333 AA.  
AC Q54147;  
DT 01-NOV-1996 (TRENBLrel. 01, Created)  
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)  
DT 01-NOV-1998 (TRENBLrel. 08, Last annotation update)  
DE SIMILAR TO E. COLI GLPT AND UHPT PROTEINS.  
GN IPGH ORFB.  
OS Shigella flexneri.  
OG Plasmid pWR100.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Shigella.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-M9OT-W, SEROTYPE 5;



RA MCMURRAY A.,  
 RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP  
 RN  
 RX  
 RX MEDLINE: 94150718.  
 RA WILSON R., AINSWORTH R., ANDERSON K., BAYNES C., BERKS M.,  
 RA BONFIELD J., BURTON J., CONNELL M., COPEY T., COOPER J., COULSON A.,  
 RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,  
 RA GARNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,  
 RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,

THIRRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,  
RA  
SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,  
RA  
PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWKNEEN R.,  
RA  
LIGHTNING J., LEVIO C., MCGOWAN R., MORILLORE B., O CALABASHAN M.,  
RA

RA WAISON A., WEINSLACK L., WILKINSON-SPOWART J., WOODHURN P.,  
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of *C.*  
elegans.";  
RL Nature 368:32-38(1994).

```

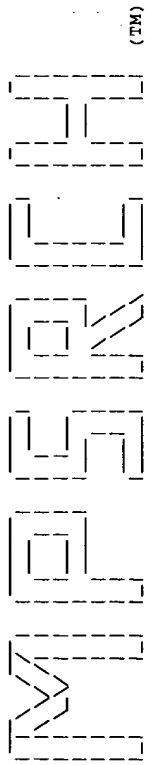
DR   EMBL; 270780; CAA94830.1; -.
DR   EMBL; 270754; CAA94830.1; JOINED.
DR   EMBL; 270754; CAA94780.1; -.
DR   EMBL; 270780; CAA94780.1; JOINED.
DR   HSS; P06601; 1FUL.
DR   PROSITE; PS00027; HOMEBOX_1; 1.
DR   PFM; PF00046; homeobox; 1.
KW   Homeobox; DNA-binding; Nuclear protein.
SQ   SEQUENCE 258 AA; 29106 MW; BDC3321C CRC32;
      Query Match 100.0%; Score 49; DB 5; Length 258;
      Best Local Similarity 36.4%; Pred. No. 9.19e+01;
      Matches 4; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
DB   166 PAAINRDGEWY 176

```

QY	2	PXXXXXXFWY	12
		1	
RESULT	5		
ID:	Q64037	PRELIMINARY;	PRT; 269 AA.
AC	Q64037;		
DT	01-NOV-1996	(TREMBlrel. 01, Created)	
DT	01-NOV-1996	(TREMBlrel. 01, Last sequence update)	
DT	01-NOV-1999	(TREMBlrel. 12, Last annotation update)	
DE	PAX-6	(FRAGMENT).	
GN	PAX-6.		
OS	Rattus norvegicus (Rat).		
OC	Eukaryota; Metazoa; Chordata;		
OC	Eutheria; Rodentia; Sciurognathi;		
RN	Muridae; Murinae; Rattus.		
	(1)		

RP SEQUENCE FROM N.A.  
RX MEDLINE; 95072652.

\*\*\*\*\*



(TM)

\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Sat Apr 15 01:43:21 2000; MasPar time 7.24 Seconds  
114.941 Million cell updates/sec

Tabular output not generated.

Title: >US-08-452-843-25  
Description: (1-12) from US08452843.pep  
Perfect Score: 49  
Sequence: 1 PXXXXXXXFWY 12

Scoring table: PAM 150  
Gap 15

Searched: 225878 seqs, 69334122 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: sprembl12  
1:sp.archaea 2:sp.bacteria 3:sp.fungi 4:sp.human  
5:sp.invertebrate 6:sp.mammal 7:sp.mhc 8:sp.organelle  
9:sp.phage 10:sp.plant 11:sp.rodent 12:sp.unclassified  
13:sp.vertebrate 14:sp.virus

Statistics: Mean 22.014; Variance 38.804; scale 0.567

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	49	100.0	142	2 Q9X957	HYPOTHETICAL 16.7 KD P	9.19e+01
2	49	100.0	180	2 Q57025	STREPTOTHRICIN-ACETYL	9.19e+01
3	49	100.0	185	2 P74640	HYPOTHETICAL 21.5 KD P	9.19e+01
4	49	100.0	258	5 Q20455	F58E6.10 PROTEIN.	9.19e+01
5	49	100.0	269	11 Q64037	PAX-6 (FRAGMENT).	9.19e+01
6	49	100.0	288	5 Q18415	SIMILAR TO PROTEIN-TYR	9.19e+01
7	49	100.0	317	5 Q16235	F07G11.1 PROTEIN.	9.19e+01
8	49	100.0	333	2 Q54147	SIMILAR TO E. COLI GLP	9.19e+01
9	49	100.0	339	1 Q9YFX4	339AA LONG HYPOTHETICA	9.19e+01
10	49	100.0	344	5 Q44901	ZK484.7 PROTEIN (EC 3.	9.19e+01
11	49	100.0	352	5 Q8XVU4	F10G8.1 PROTEIN (EC 3.	9.19e+01
12	49	100.0	356	10 Q9XJ40	CHLOROPHYLL B SYNTHASE	9.19e+01
13	49	100.0	367	11 Q70169	TESTICULAR SERINE PROT	9.19e+01
14	49	100.0	372	2 P74455	HIGH-AFFINITY BRANCHED	9.19e+01
15	49	100.0	407	5 Q44371	UNC-129.	9.19e+01
16	49	100.0	429	10 Q23838	S GLYCOPROTEIN (FRAGME	9.19e+01
17	49	100.0	431	10 Q23847	S GLYCOPROTEIN (FRAGME	9.19e+01
18	49	100.0	431	10 Q23849	S GLYCOPROTEIN (FRAGME	9.19e+01
19	49	100.0	451	10 Q40098	SECRETED GLYCOPROTEIN	9.19e+01
20	49	100.0	451	10 Q40099	SECRETED GLYCOPROTEIN	9.19e+01

21	49	100.0	507	3 Q00079	ACID PHOSPHATASE (FRAG	9.19e+01
22	49	100.0	536	10 Q9XJ37	CHLOROPHYLL B SYNTHASE	9.19e+01
23	49	100.0	563	5 Q18814	C53D6.2 PROTEIN.	9.19e+01
24	49	100.0	576	14 Q85433	CAPSID PROTEIN VP60.	9.19e+01
25	49	100.0	610	14 P87554	TERMINAL PROTEIN PRECU	9.19e+01
26	49	100.0	614	3 Q12546	ACID PHOSPHATASE PRECU	9.19e+01
27	49	100.0	655	2 Q9X0F7	OLIGOPEPTIDE ABC TRANS	9.19e+01
28	49	100.0	660	2 Q9XR2	OLIGOPEPTIDE ABC TRANS	9.19e+01
29	49	100.0	685	3 Q9Y8H3	BRANCHING ENZYME.	9.19e+01
30	49	100.0	737	10 Q80716	F14M4.8 PROTEIN.	9.19e+01
31	49	100.0	853	10 Q40096	RECEPTOR PROTEIN KINAS	9.19e+01
32	49	100.0	1307	3 Q43138	MULTIDRUG RESISTANCE P	9.19e+01
33	49	100.0	1613	1 Q58907	PUTATIVE REVERSE GYRAS	9.19e+01
34	49	100.0	1813	5 Q17665	UNC-13 PROTEIN.	9.19e+01
35	45	91.8	102	1 Q9YB5	102AA LONG HYPOTHETICA	3.52e+02
36	45	91.8	163	5 Q96584	LONG-WAVELENGTH RHODOP	3.52e+02
37	45	91.8	332	2 Q32480	FERRIC IRON BINDING PR	3.52e+02
38	45	91.8	332	2 Q9ZFR1	PUTATIVE PHATHALATE PE	3.52e+02
39	45	91.8	374	14 Q89634	GLYCOPROTEIN GD.	3.52e+02
40	45	91.8	398	14 Q9YS31	GLYCOPROTEIN GP50.	3.52e+02
41	45	91.8	404	14 Q92230	GLYCOPROTEIN GD PRECUR	3.52e+02
42	45	91.8	552	2 Q44528	PUTATIVE ALPHA-GLUCAN	3.52e+02
43	45	91.8	825	2 Q59319	ALPHA-DEXTRIN 6-GLUCAN	3.52e+02
44	45	91.8	901	10 Q80462	PUTATIVE GUANINE NUCLE	3.52e+02
45	45	91.8	1232	2 Q06559	NARG.	3.52e+02

ALIGNMENTS

RESULT	1	PRELIMINARY;	PRT;	142 AA.
ID	Q9X957;			
AC	Q9X957;			
DT	01-NOV-1999 (Tremblrel. 12, Created)			
DT	01-NOV-1999 (Tremblrel. 12, Last sequence update)			
DE	01-NOV-1999 (Tremblrel. 12, Last annotation update)			
DE	HYPOTHETICAL 16.7 KD PROTEIN (FRAGMENT).			
OS	Synechococcus elongatus.			
OC	Bacteria; Cyanobacteria; Chroococcales; Synechococcus.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	SCHNEIDER D., ALTENFELD U., THOMAS H., SCHRADER S., MUEHLENHOFF U., ROEGNER M.;			
RT	"Cloning and sequencing of two operons encoding the four major subunits of the cytochrome b6f complex of the cyanobacterium Synechococcus elongatus.";			
RL	Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; AJ243707; CAB46751.1; -			
KW	Hypothetical protein.			
FT	NON_TER 1			
SQ	SEQUENCE 142 AA; 16726 MW; CDB88BC6 CRC32;			

Query Match 100.0%; Score 49; DB 2; Length 142;  
Best Local Similarity 36.4%; Pred. No. 9.19e+01;  
Matches 4; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Db	7 PKIVLVNAFWY 17
QY	2 PXXXXXXXFWY 12

RESULT 2  
ID Q57025; PRELIMINARY; PRT; 180 AA.

AC	Q57025;
DT	01-NOV-1996 (Tremblrel. 01, Created)
DT	01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT	01-JAN-1999 (Tremblrel. 09, Last annotation update)
DE	STREPTOTHRICIN-ACETYLTRANSFERASE.
GN	SAT3.
OS	Escherichia coli, and unidentified.
OG	Plasmid pIE1107, and plasmid pIE539.
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.
RN	[1]

RL Gene 71:339-348(1988).  
CC -1- CATALYTIC ACTIVITY: AN ORTHOPHOSPHORIC MONOESTER + H(2)O -> AN  
CC ALCOHOL + ORTHOPHOSPHATE.  
CC -1- SUBUNIT: MONOMER.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; M23540; AAA32700.1; -  
DR PIR; J0386; J0386.  
KW Hydrolase; Glycoprotein; Signal.  
FT SIGNAL 1 20 POTENTIAL.  
FT CHAIN 21 436 PHOSPHATE-REPRESSIBLE ACID PHOSPHATASE.  
FT CARBOHYD 227 227 POTENTIAL.  
FT CARBOHYD 283 283 POTENTIAL.  
FT CARBOHYD 304 304 POTENTIAL.  
SQ SEQUENCE 436 AA; 47995 MW; 59126ADD CRC32;

Query Match 100.0%; Score 49; DB 1; Length 436;  
Best Local Similarity 36.4%; Pred. No. 4.50e+01;  
Matches 4; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Db 240 PETGGVGNFWY 250  
QY 2 PXXXXXXFWY 12

RESULT 15  
ID PCNB HAEIN STANDARD; PRT; 452 AA.  
AC P44439.  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 01-OCT-1996 (Rel. 34, Last annotation update)  
DE PROBABLE POLY(A) POLYMERASE (EC 2.7.7.19) (PAP).  
GN PCNB OR HI0063.  
OS Haemophilus influenzae.  
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
OC Haemophilus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-RD / KW20;  
RX MEDLINE; 95350630.  
RA FLEISCHMANN R.D., ADAMS M.D., WHITE O., CLAYTON R.A., KIRKNESS E.F.,  
RA KERLAVAGE A.R., BULT C.J., TOMB J.-F., DOUGHERTY B.A., MERRICK J.M.,  
RA MCKENNEY K., SUTTON G., FITZHUGH W., FIELDS C.A., GOCAYNE J.D.,  
RA SCOTT J.D., SHIRLEY R., LIU L.-I., GLODEK A., KELLEY J.M.,  
RA WEIDMAN J.F., PHILLIPS C.A., SPRIGGS T., HEDBLUM E., COTTON M.D.,  
RA UTTERBACK T.R., HANNA M.C., NGUYEN D.T., SAUDEK D.M., BRANDON R.C.,  
RA FINE L.D., FRITCHMAN J.L., FUHRMANN J.L., GEOGHAGEN N.S.M.,  
RA GNEHM C.L., McDONALD L.A., SMALL K.V., FRASER C.M., SMITH H.O.,  
RA VENTER J.C.;  
RT "Whole-genome random sequencing and assembly of Haemophilus  
RT influenzae Rd."  
RL Science 269:496-512(1995).  
CC -1- FUNCTION: POLYMERASE THAT CREATES THE 3' POLY(A) TAIL FOUND IN  
CC SOME MRNA'S (BY SIMILARITY).  
CC -1- CATALYTIC ACTIVITY: N ATP + (NUCLEOTIDE)(M) = N PYROPHOSPHATE +  
CC (NUCLEOTIDE)(M+N).  
CC -1- SUBUNIT: MONOMER (BY SIMILARITY).  
CC -1- SIMILARITY: BELONGS TO THE TRNA NUCLEOTIDYLTRANSFERASE / POLY(A)  
CC POLYMERASE FAMILY.  
CC -----

This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).  
CC -----

DR EMBL; U32691; AAC21741.1; -  
DR TIGR; HI0063; -  
KW mRNA processing; Transferase; Transcription; RNA-binding.  
FT ACT\_SITE 68 68 BY SIMILARITY.  
FT ACT\_SITE 70 70 BY SIMILARITY.  
FT ACT\_SITE 150 150 BY SIMILARITY.  
SQ SEQUENCE 452 AA; 52697 MW; BC80CFC2 CRC32;

Query Match 100.0%; Score 49; DB 1; Length 452;  
Best Local Similarity 36.4%; Pred. No. 4.50e+01;  
Matches 4; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Db 302 PAFLFAAFFWY 312  
QY 2 PXXXXXXFWY 12

Search completed: Sat Apr 15 01:43:03 2000  
Job time : 40 secs.

```
FT HELIX 341 352
FT STRAND 354 355
FT TURN 360 361
FT HELIX 362 377
FT TURN 378 379
FT HELIX 383 394
SQ SEQUENCE 396 AA; 43387 MW; 93EF40AE CRC32;

Query Match 100.0%; Score 49; DB 1; Length 396;
Best Local Similarity 36.4%; Pred. No. 4.50e+01;
Matches 4; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Db 357 PNIPQMSAFWY 367
QY 2 PXXXXXXXFWY 12

RESULT 12
ID MALE-ENTAE STANDARD; PRT; 396 AA.
AC P18815;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE MALTOSE-BINDING PERIPLASMIC PROTEIN PRECURSOR (MALTODEXTRIN-BINDING
DE PROTEIN) (MWPB).
GN MALE.
OS Enterobacter aerogenes (Aerobacter aerogenes).
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Enterobacter.
[1]
RN SEQUENCE FROM N.A.
RX MEDLINE; 89384443.
RA DAHL M.K., FRANCOZ E., SAURIN W., BOOS W., MANSON M.D., HOFNUNG M.;
RT "Comparison of sequences from the malB regions of Salmonella
RT typhimurium and Enterobacter aerogenes with Escherichia coli K12: a
RL potential new regulatory site in the interperonic region.";
RL Mol. Gen. Genet. 218:199-207(1989).
CC -!- FUNCTION: MALE IS INVOLVED IN THE HIGH-AFFINITY MALTOSE MEMBRANE
CC TRANSPORT SYSTEM. INITIAL RECEPTOR FOR THE ACTIVE TRANSPORT OF AND
CC CHEMOTAXIS TOWARD MALTOLOGOSACCHARIDES.
CC -!- SUBCELLULAR LOCATION: PERIPLASMIC.
CC -!- SIMILARITY: BELONGS TO THE BACTERIAL EXTRACELLULAR SOLUTE-BINDING
CC PROTEIN FAMILY 1.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; X54292; CAA38189.1; .
DR PIR; S05331; S05331.
DR PIR; S20603; S20603.
DR HSP; P02928; IMDP.
DR STYGENE; SG10212; MALE.
DR PROSITE; PS01037; SBP_BACTERIAL_1; 1.
DR PFAM; PF01547; SBP_bacterial_1; 1.
KW Transport; Sugar transport; Periplasmic; Signal.
FT SIGNAL 1 26
FT CHAIN 27 396 MALTOSE-BINDING PERIPLASMIC PROTEIN.
SQ SEQUENCE 396 AA; 43136 MW; 3C20F290 CRC32;

Query Match 100.0%; Score 49; DB 1; Length 396;
Best Local Similarity 36.4%; Pred. No. 4.50e+01;
Matches 4; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Db 357 PNIPQMSAFWY 367
QY 2 PXXXXXXXFWY 12

RESULT 13
ID MALE-SALTY STANDARD; PRT; 396 AA.
AC P19576;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE MALTOSE-BINDING PERIPLASMIC PROTEIN PRECURSOR (MALTODEXTRIN-BINDING
DE PROTEIN) (MWPB).
GN MALE.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
```

```
RN SEQUENCE FROM N.A.
RP MEDLINE; 89384443.
RA DAHL M.K., FRANCOZ E., SAURIN W., BOOS W., MANSON M.D., HOFNUNG M.;
RT "Comparison of sequences from the malB regions of Salmonella
RT typhimurium and Enterobacter aerogenes with Escherichia coli K12: a
RL potential new regulatory site in the interperonic region.";
RL Mol. Gen. Genet. 218:199-207(1989).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-LT2;
RX MEDLINE; 92110387.
RA SCHNEIDER E., FRANCOZ E., DASSA E.;
RT "Completion of the nucleotide sequence of the 'maltose B' region in
RT Salmonella typhimurium: the high conservation of the malM gene
RT suggests a selected physiological role for its product.";
RL Biochim. Biophys. Acta 1129:223-227(1992).
CC -!- FUNCTION: MALE IS INVOLVED IN THE HIGH-AFFINITY MALTOSE MEMBRANE
CC TRANSPORT SYSTEM. INITIAL RECEPTOR FOR THE ACTIVE TRANSPORT OF AND
CC CHEMOTAXIS TOWARD MALTOLOGOSACCHARIDES.
CC -!- SUBCELLULAR LOCATION: PERIPLASMIC.
CC -!- SIMILARITY: BELONGS TO THE BACTERIAL EXTRACELLULAR SOLUTE-BINDING
CC PROTEIN FAMILY 1.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; X54292; CAA38189.1; .
DR PIR; S05331; S05331.
DR PIR; S20603; S20603.
DR HSP; P02928; IMDP.
DR STYGENE; SG10212; MALE.
DR PROSITE; PS01037; SBP_BACTERIAL_1; 1.
DR PFAM; PF01547; SBP_bacterial_1; 1.
KW Transport; Sugar transport; Periplasmic; Signal.
FT SIGNAL 1 26
FT CHAIN 27 396 MALTOSE-BINDING PERIPLASMIC PROTEIN.
SQ SEQUENCE 396 AA; 43152 MW; 5A4E61DD CRC32;

Query Match 100.0%; Score 49; DB 1; Length 396;
Best Local Similarity 36.4%; Pred. No. 4.50e+01;
Matches 4; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Db 357 PNIPQMSAFWY 367
QY 2 PXXXXXXXFWY 12

RESULT 14
ID PPAL-ASPNG STANDARD; PRT; 436 AA.
AC P20584;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 01-OCT-1994 (Rel. 30, Last annotation update)
DE PHOSPHATE-REPRESSIBLE ACID PHOSPHATASE PRECURSOR (EC 3.1.3.2) (ACID
DE PHOSPHATASE PII).
GN PACA OR APHA.
OS Aspergillus niger.
OC Eukaryota; Fungi; Ascomycota; Euascomycetes; Plectomycetes;
OC Eurotiata; Trichocomaceae; mitosporic trichocomaceae; Aspergillus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 46951;
RX MEDLINE; 89138010.
RA MACRAE W.D., BUXTON F.P., SIBLEY S., GARVEN S., GWYNNE D.I.,
RA DAVIES R.W., ARST H.N. JR.;
RT "A phosphate-repressible acid phosphatase gene from Aspergillus
RT niger: its cloning, sequencing and transcriptional analysis.";
RT
```

RP SEQUENCE OF 1-32 FROM N.A.  
RX MEDLINE: 83111968.  
RA BEDOUELLE H., SCHMEISSNER U., HOFNUNG M., ROSENBERG M.;  
RT "Promoters of the malefeg and malk-lamb operons in Escherichia coli  
RT K12.";  
RL J. Mol. Biol. 161:519-531(1982).  
RN [5]  
RP SEQUENCE OF 392-396 FROM N.A.  
RX MEDLINE: 84289514.  
RA FROSHAUER S., BECKWITH J.;  
RT "The nucleotide sequence of the gene for malP protein, an inner  
RT membrane component of the maltose transport system of Escherichia  
RT coli. Repeated DNA sequences are found in the male-malf  
RT intercalon region.";  
RL J. Biol. Chem. 259:10896-10903(1984).  
RN [6]  
RP SEQUENCE OF 27-38.  
RC STRAIN-K12 / EMG2;  
RX MEDLINE: 97443975.  
RA LINK A.J., ROBISON K., CHURCH G.M.;  
RT "Comparing the predicted and observed properties of proteins encoded  
RT in the genome of Escherichia coli K-12.";  
RL Electrophoresis 18:1259-1313(1997).  
RN [7]  
RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS).  
RX MEDLINE: 91161615.  
RA SPURLINO J.C., LU G.-Y., QUICHO F.A.;  
RT "The 2.3-A resolution structure of the maltose-or  
RT maltodextrin-binding protein, a primary receptor of bacterial active  
RT transport and chemotaxis.";  
RL J. Biol. Chem. 266:5202-5219(1991).  
RN [8]  
RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS).  
RX MEDLINE: 93041761.  
RA SHARFF A.J., RODSETH L.E., SPURLINO J.C., QUICHO F.A.;  
RT "Crystallographic evidence of a large ligand-induced hinge-twist  
RT motion between the two domains of the maltodextrin binding protein  
RT involved in active transport and chemotaxis.";  
RL Biochemistry 31:10657-10663(1992).  
RN [9]  
RP X-RAY CRYSTALLOGRAPHY (1.7 ANGSTROMS).  
RX MEDLINE: 97454787.  
RA QUICHO F.A., SPURLINO J.C., RODSETH L.E.;  
RT "Extensive features of tight oligosaccharide binding revealed in  
RT high-resolution structures of the maltodextrin  
RT transport/chemosensory receptor.";  
RL Structure 5:997-1015(1997).  
CC -1- FUNCTION: INVOLVED IN THE HIGH-AFFINITY MALTOSE MEMBRANE TRANSPORT  
CC SYSTEM MALEFGK. INITIAL RECEPTOR FOR THE ACTIVE TRANSPORT OF AND  
CC CHEMOTAXIS TOWARD MALTOLOGOSACCHARIDES.  
CC -1- SUBCELLULAR LOCATION: PERIPLASMIC.  
CC -1- SIMILARITY: BELONGS TO THE BACTERIAL EXTRACELLULAR SOLUTE-BINDING  
CC PROTEIN FAMILY 1.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: V00303; CAA23581.1; -  
DR EMBL: U00006; AAC43128.1; -  
DR EMBL: AF000476; AAC77004.1; -  
DR EMBL: J01648; CAB23257.1; ALT\_INIT.  
DR EMBL: M16181; AAA24102.1; -  
DR EMBL: M12635; AAA24123.1; -  
DR EMBL: M12647; AAA24135.1; -  
DR EMBL: M12650; AAA24138.1; -  
DR EMBL: M12653; AAA24115.1; -  
DR PIR: A03428; JGECW.  
PDB: 3MBP; 24-DEC-97.

DR PDB: 4MBP; 25-FEB-98.  
DR PDB: IOMP; 31-JAN-94.  
DR PDB: IDMB; 31-OCT-93.  
DR PDB: IMDP; 01-NOV-94.  
DR PDB: IMDQ; 01-NOV-94.  
DR PDB: IMPB; 15-OCT-95.  
DR PDB: IMPC; 15-OCT-95.  
DR PDB: IMPD; 15-OCT-95.  
DR PDB: IANF; 24-DEC-97.  
DR PDB: ITUD; 05-JUN-97.  
DR PDB: IA7L; 17-JUN-98.  
DR SWISS-2DPAGE; P02928; COLI.  
DR ECOGENE; EG10554; MALE.  
DR PROSITE; PS01037; SBP\_BACTERIAL\_1; 1.  
DR PFAM; PF01547; SBP\_BACTERIAL\_1; 1.  
KW Transport; Sugar transport; Periplasmic; Signal; 3D-structure.  
FT SIGNAL 1 26  
FT CHAIN 27 396  
FT CONFLICT 36 36  
FT CONFLICT 46 46  
FT STRAND 33 36  
FT TURN 39 40  
FT TURN 43 57  
FT STRAND 61 64  
FT TURN 67 68  
FT TURN 69 77  
FT TURN 78 80  
FT STRAND 85 89  
FT TURN 90 98  
FT TURN 99 100  
FT STRAND 102 102  
FT TURN 109 112  
FT TURN 113 114  
FT TURN 115 115  
FT TURN 117 123  
FT STRAND 124 125  
FT TURN 126 127  
FT STRAND 128 129  
FT TURN 132 137  
FT STRAND 140 144  
FT TURN 145 147  
FT STRAND 154 154  
FT TURN 156 157  
FT TURN 158 166  
FT TURN 167 169  
FT TURN 171 173  
FT TURN 180 189  
FT TURN 190 191  
FT STRAND 193 197  
FT STRAND 202 208  
FT TURN 212 226  
FT TURN 227 228  
FT TURN 232 233  
FT TURN 236 244  
FT TURN 245 246  
FT STRAND 248 253  
FT TURN 255 264  
FT TURN 265 265  
FT STRAND 268 271  
FT TURN 275 276  
FT TURN 277 278  
FT STRAND 279 279  
FT STRAND 284 293  
FT TURN 294 295  
FT TURN 297 298  
FT TURN 299 308  
FT TURN 309 310  
FT TURN 313 322  
FT STRAND 327 328  
FT STRAND 330 330  
FT TURN 331 337  
FT TURN 338 339

MALTOSE-BINDING PERIPLASMIC PROTEIN.  
W -> A (IN REF. 6).  
L -> H (IN REF. 3).

DR PROSITE; PS00237; G\_PROTEIN\_RECEPTOR; 1.  
DR PFAM; PF00001; 7tm\_1; 1.  
KW G-protein coupled receptor; Transmembrane; Glycoprotein;  
KW Multigene family; Lipoprotein; Palmitate.  
FT DOMAIN 1 20 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 21 41 1 (POTENTIAL).  
FT DOMAIN 42 57 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 58 78 2 (POTENTIAL).  
FT DOMAIN 79 96 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 97 118 3 (POTENTIAL).  
FT DOMAIN 119 138 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 139 159 4 (POTENTIAL).  
FT DOMAIN 160 184 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 185 205 5 (POTENTIAL).  
FT DOMAIN 206 236 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 237 257 6 (POTENTIAL).  
FT DOMAIN 258 259 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 260 280 7 (POTENTIAL).  
FT DOMAIN 281 370 CYTOPLASMIC (POTENTIAL).  
FT CARBOHYD 6 6 POTENTIAL.  
FT DISULFID 95 172 BY SIMILARITY.  
FT LIPID 308 308 PALMITATE (BY SIMILARITY).  
FT DOMAIN 218 225 POLY-ALA.  
FT DOMAIN 313 318 POLY-HIS.  
SQ SEQUENCE 370 AA; 40390 MW; 62754539 CRC32;

Query Match 100.0%; Score 49; DB 1; Length 370;  
Best Local Similarity 36.4%; Pred.No. 4.50e+01;  
Matches 4; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Db 250 PHHALIICFWY 260  
QY 2 PXXXXXXFWY 12

RESULT 10  
ID GALT\_RAT STANDARD; PRT; 370 AA.  
AC O88626; O54914;  
DT 15-JUL-1999 (Rel. 38, Created)  
DT 13-JUL-1999 (Rel. 38, Last sequence update)  
DT 13-DEC-1999 (Rel. 39, Last annotation update)  
DE GALANIN RECEPTOR TYPE 3 (GALR3) (GALR3).  
GN GALR3 OR GALR3.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;  
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-SPRAGUE-DAWLEY; TISSUE-HYPOTHALAMUS;  
RA WATERS S.M., KRAUSE J.E.;  
RT "Rat galanin receptor type 3 (GALR3) coding region.";  
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA WANG S., HE C., HASHEMI T., BAYNE M.;  
RT "Molecular cloning of the rat galanin receptor type 3.";  
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: RECEPTOR FOR THE HORMONE GALANIN.  
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.  
CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AF079844; AAC34590.1; -  
DR EMBL; AF031522; AAC26145.1; -  
DR GCRDB; GCR\_2508; -  
DR PROSITE; PS00237; G\_PROTEIN\_RECEPTOR; 1.

DR PFAM; PF00001; 7tm\_1; 1.  
KW G-protein coupled receptor; Transmembrane; Glycoprotein;  
KW Multigene family; Lipoprotein; Palmitate.  
FT DOMAIN 1 20 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 21 41 1 (POTENTIAL).  
FT DOMAIN 42 57 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 58 78 2 (POTENTIAL).  
FT DOMAIN 79 96 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 97 118 3 (POTENTIAL).  
FT DOMAIN 119 138 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 139 159 4 (POTENTIAL).  
FT DOMAIN 160 184 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 185 205 5 (POTENTIAL).  
FT DOMAIN 206 236 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 237 257 6 (POTENTIAL).  
FT DOMAIN 258 259 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 260 280 7 (POTENTIAL).  
FT DOMAIN 281 370 CYTOPLASMIC (POTENTIAL).  
FT CARBOHYD 6 6 POTENTIAL.  
FT DISULFID 95 172 BY SIMILARITY.  
FT LIPID 308 308 PALMITATE (BY SIMILARITY).  
FT DOMAIN 218 225 POLY-ALA.  
FT CONFLICT 127 127 P -> Q (IN REF. 2).  
FT CONFLICT 183 183 A -> R (IN REF. 2).  
FT CONFLICT 311 311 R -> C (IN REF. 2).  
SQ SEQUENCE 370 AA; 40410 MW; 1EBF1E77 CRC32;

Query Match 100.0%; Score 49; DB 1; Length 370;  
Best Local Similarity 36.4%; Pred.No. 4.50e+01;  
Matches 4; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Db 250 PHHALIICFWY 260  
QY 2 PXXXXXXFWY 12

RESULT 11  
ID MALE\_ECOLI STANDARD; PRT; 396 AA.  
AC P02928;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 21-JUL-1986 (Rel. 01, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE MALTOSE-BINDING PERIPLASMIC PROTEIN PRECURSOR (MALTODEXTRIN-BINDING  
DE PROTEIN) (MMPB).  
GN MALE  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Escherichia.  
RN [1]  
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
RC STRAIN-K12;  
RX MEDLINE; 84289460.  
RA DUPLAY P., BEDOUELLE H., FOWLER A., ZABIN I., SAURIN W., HOFNUNG M.;  
RT "Sequences of the male gene and of its product, the maltose-binding  
RL protein of Escherichia coli K12.";  
RL J. Biol. Chem. 259:10606-10613(1984).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-K12 / MG1655;  
RX MEDLINE; 94089392.  
RA BLATTNER F.R., BURLAND V.D., PLUNKETT G. III, SOFIA H.J.,  
RA DANIELS D.L.;  
RT "Analysis of the Escherichia coli genome. IV. DNA sequence of the  
RT region from 89.2 to 92.8 minutes.";  
RL Nucleic Acids Res. 21:5408-5417(1993).  
RN [3]  
RP SEQUENCE OF 1-48 FROM N.A.  
RX MEDLINE; 82219202.  
RA BEDOUELLE H., HOFNUNG M.;  
RT "A DNA sequence containing the control regions of the maleFG and  
RL malk-lamb operons in Escherichia coli K12.";  
RL Mol. Gen. Genet. 185:82-87(1982).  
RN [4]

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; 267753; CRA91611.1; -  
 DR EMBL; 267753; CRA91667.1; -  
 KW Chloroplast; Hypothetical protein.  
 SQ SEQUENCE 355 AA; 41643 MW; 994A564F CRC32;

Query Match 100.0%; Score 49; DB 1; Length 355;  
 Best Local Similarity 36.4%; Pred. No. 4.50e+01;  
 Matches 4; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Db 254 RPDRLREFWY 264  
 QY 2 PXXXXXXFWY 12

RESULT 8  
 ID GALT HUMAN STANDARD; PRT; 368 AA.  
 AC O60755;  
 DT 15-JUL-1999 (Rel. 38, Created)  
 DT 15-JUL-1999 (Rel. 38, Last sequence update)  
 DE GALANIN RECEPTOR TYPE 3 (GAL3-R) (GALR3).  
 GN GALNR3 OR GALR3.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 98389766.  
 RA SMITH K.E., WALKER M.W., ARTYMYSHYN R., BARD J., BOROWSKY B.,  
 RA TAMM J.A., YAO W.-J., VAYSE P.J.-J., BRANCHER T.A., GERALD C.,  
 RA JONES K.A.;  
 RT "Cloned human and rat galanin GALR3 receptors: pharmacology and  
 RT activation of G-protein inwardly rectifying K<sup>+</sup> channels.";  
 RL J. Biol. Chem. 273:23321-23326(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC BENNETT M.M., LESCOE M.K., GALLIPOLI P.Z., RAMABHADRAN T.V.;  
 RT "Homologue of the human galanin 2 receptor gene isolated from a human  
 RT uterus cDNA library.";  
 RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 99048961.  
 RA KOLAKOWSKI L.F. JR., O'NEILL G.P., HOWARD A.D., BROUSSARD S.R.,  
 RA SULLIVAN K.A., FEIGNER S.D., SAWZDARGO M., NGUYEN T., KARGMAN S.,  
 RA SHAO L.-L., HRENIUK D.L., TAN C.P., EVANS J., ABRAMOVITZ M.,  
 RA CHATEAUNEUF A., COULOMBE N., NG G., JOHNSON M.P., THARIAN A.,  
 RA KHOSHBOUEI H., GEORGE S.R., SMITH R.G., O'DOWD B.F.;  
 RT "Molecular characterization and expression of cloned human galanin  
 RT receptors GALR2 and GALR3.";  
 RL J. Neurochem. 71:2239-2251(1998).  
 CC -1- FUNCTION: RECEPTOR FOR THE HORMONE GALANIN.  
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.  
 CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; AF073799; AAC35944.1; -  
 DR EMBL; AF067733; AAC18860.1; -

DR MM; 603692; -  
 DR PROSITE; PS00237; G\_PROTEIN\_RECEPTOR; 1.  
 PFAM; PF00001; 7tm.1; 1.  
 KW G-protein coupled receptor; Transmembrane; Glycoprotein;  
 KW Multigene family; Lipoprotein; Palmitate.  
 FT DOMAIN 1 20  
 FT TRANSSEM 21 41  
 FT DOMAIN 42 57  
 FT TRANSSEM 58 78  
 FT DOMAIN 79 96  
 FT TRANSSEM 97 118  
 FT DOMAIN 119 138  
 FT TRANSSEM 139 159  
 FT DOMAIN 160 184  
 FT TRANSSEM 185 205  
 FT DOMAIN 206 236  
 FT TRANSSEM 237 257  
 FT DOMAIN 258 280  
 FT TRANSSEM 281 368  
 FT CARBOHYD 6  
 FT DISULFID 95 172  
 FT LIPID 308 308  
 FT DOMAIN 218 225  
 FT DOMAIN 310 318  
 SQ SEQUENCE 368 AA; 39573 MW; 2DF74618 CRC32;

Query Match 100.0%; Score 49; DB 1; Length 368;  
 Best Local Similarity 36.4%; Pred. No. 4.50e+01;  
 Matches 4; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Db 250 PHHALILCFWY 260  
 QY 2 PXXXXXXFWY 12

RESULT 9  
 ID GALT\_MOUSE STANDARD; PRT; 370 AA.  
 AC O88853;  
 DT 15-JUL-1999 (Rel. 38, Created)  
 DT 15-JUL-1999 (Rel. 38, Last sequence update)  
 DE GALANIN RECEPTOR TYPE 3 (GAL3-R) (GALR3).  
 GN GALNR3 OR GALR3.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-129/SV;  
 RX MEDLINE; 99048961.  
 RA KOLAKOWSKI L.F. JR., O'NEILL G.P., HOWARD A.D., BROUSSARD S.R.,  
 RA SULLIVAN K.A., FEIGNER S.D., SAWZDARGO M., NGUYEN T., KARGMAN S.,  
 RA SHAO L.-L., HRENIUK D.L., TAN C.P., EVANS J., ABRAMOVITZ M.,  
 RA CHATEAUNEUF A., COULOMBE N., NG G., JOHNSON M.P., THARIAN A.,  
 RA KHOSHBOUEI H., GEORGE S.R., SMITH R.G., O'DOWD B.F.;  
 RT "Molecular characterization and expression of cloned human galanin  
 RT receptors GALR2 and GALR3.";  
 RL J. Neurochem. 71:2239-2251(1998).  
 CC -1- FUNCTION: RECEPTOR FOR THE HORMONE GALANIN.  
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.  
 CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; AF042783; AAC36588.1; -  
 DR MGD; MGI:1329003; GALR3.

RX MEDLINE; 94237472.  
RA SIKSNYS V., ZAREKRAJA N., VAISVILA R., TIMINSKAS A., STAKENAS P.,  
RA BUTKUS V., JANULAITIS A.;  
RT "CAATG-specific restriction-modification muni genes from Mycoplasma:  
RT sequence similarities between R.Muni and R.ECORI";  
RL Gene 142:1-8(1994).  
CC -1- CATALYTIC ACTIVITY: RECOGNIZES THE DOUBLE-STRANDED SEQUENCE CAATG  
CC AND CLEAVES AFTER C-1.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; X76192; CAA53788.1; --  
DR REBASE; RB02037; Muni.  
KW Hydrolyase; Endonuclease; Nuclease; Restriction system.  
SQ SEQUENCE 202 AA; 23389 MW; D3B4F3B7 CRC32;  
  
Query Match 100.0%; Score 49; DB 1; Length 202;  
Best Local Similarity 36.4%; Pred. No. 4.50e+01;  
Matches 4; Conservative 0; Mismatches 7; Indels 0; Gaps 0;  
  
Db 160 PKRVREITFWY 170  
QY 2 PXXXXXXFWY 12  
|||  
  
RESULT 6  
ID CCAD\_MOUSE STANDARD; PRT; 281 AA.  
AC Q99246;  
DT 15-JUL-1999 (Rel. 38, Created)  
DT 15-JUL-1999 (Rel. 38, Last sequence update)  
DE VOLTAGE-DEPENDENT L-TYPE CALCIUM CHANNEL ALPHA-1D SUBUNIT (CALCIUM  
DE CHANNEL, L TYPE, ALPHA-1 POLYPEPTIDE ISOFORM 2) (FRAGMENT).  
GN CACNALD OR CACNL1A2 OR CCHL1A2 OR CACH3 OR CACNA4.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
[1]  
SEQUENCE FROM N.A. (ISOFORM CACH3B).  
RP STRAIN=ICR; TISSUE=OVARY;  
RX MEDLINE; 91056091;  
RA PEREZ-REYES E., WEI X., CASTELLANO A., BIRNBAUMER L.;  
RT "Molecular diversity of L-type calcium channels. Evidence for  
RT alternative splicing of the transcripts of three non-allelic genes";  
RL J. Biol. Chem. 265:20430-20436(1990).  
CC -1- FUNCTION: VOLTAGE-SENSITIVE CALCIUM CHANNELS (VSCC) MEDIATE THE  
CC ENTRY OF CALCIUM IONS INTO EXCITABLE CELLS AND ARE ALSO INVOLVED  
CC IN A VARIETY OF CALCIUM-DEPENDENT PROCESSES, INCLUDING MUSCLE  
CC CONTRACTION, HORMONE OR NEUROTRANSMITTER RELEASE, GENE EXPRESSION,  
CC CELL MOTILITY, CELL DIVISION AND CELL DEATH. THE ISOFORM ALPHA-1D  
CC GIVES RISE TO L-TYPE CALCIUM CURRENTS. LONG-LASTING (L-TYPE)  
CC CALCIUM CHANNELS BELONG TO THE "HIGH-VOLTAGE ACTIVATED" (HVA)  
CC GROUP. THEY ARE BLOCKED BY DIHYDROPYRIDINES (DHP),  
CC PHENTOLAMINES, BENZOTHAZEPINES, AND BY OMEGA-AGATOXIN-11IA  
CC (OMEGA-AGA-11IA). THEY ARE HOWEVER INSENSITIVE TO OMEGA-CONOTOXIN-  
CC GVIA (OMEGA-CTX-GVIA) AND OMEGA-AGATOXIN-IVA (OMEGA-AGA-IVA).  
CC -1- SUBUNIT: VOLTAGE-DEPENDENT CALCIUM CHANNELS ARE MULTISUBUNIT  
CC COMPLEXES, CONSISTING OF ALPHA-1, ALPHA-2, BETA AND DELTA SUBUNITS  
CC IN A 1:1:1:1 RATIO. THE CHANNEL ACTIVITY IS DIRECTED BY THE FOR-  
CC FORMING AND VOLTAGE-SENSITIVE ALPHA-1 SUBUNIT. IN MANY CASES, THIS  
CC SUBUNIT IS SUFFICIENT TO GENERATE VOLTAGE-SENSITIVE CALCIUM  
CC CHANNEL ACTIVITY. THE AUXILIARY SUBUNITS BETA AND ALPHA-2/DELTA  
CC LINKED BY A DISULFIDE BRIDGE REGULATE THE CHANNEL ACTIVITY.  
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.  
CC -1- ALTERNATIVE PRODUCTS: MULTIPLE ISOFORMS ARE PRODUCED BY  
CC ALTERNATIVE SPLICING (PROBABLE). THE SEQUENCE SHOWN HERE IS THAT  
CC OF CACH3B.

CC -1- DOMAIN: EACH OF THE FOUR INTERNAL REPEATS CONTAINS FIVE  
CC HYDROPHOBIC TRANSMEMBRANE SEGMENTS (S1, S2, S3, S5, S6) AND ONE  
CC POSITIVELY CHARGED TRANSMEMBRANE SEGMENT (S4). S4 SEGMENTS  
CC PROBABLY REPRESENT THE VOLTAGE-SENSOR AND ARE CHARACTERIZED BY A  
CC SERIES OF POSITIVELY CHARGED AMINO ACIDS AT EVERY THIRD POSITION.  
CC -1- SIMILARITY: BELONGS TO THE CALCIUM CHANNEL ALPHA-1 SUBUNITS  
CC FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; M57975; AAA63292.1; --  
DR MGD; MGI:88293; CACNALD.  
DR PFAM; PF00520; ion\_trans. 1.  
KW Ionic channel; Transmembrane; Ion transport; Voltage-gated channel;  
KW Calcium channel; Glycoprotein; Repeat; Multigene family;  
KW Calcium-binding; Alternative splicing.  
FT NON\_TER 1 1  
FT REPEAT 45 >281 IV  
FT DOMAIN 2 58 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 59 77 S1 OF REPEAT IV (POTENTIAL).  
FT DOMAIN 78 92 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 93 112 S2 OF REPEAT IV (POTENTIAL).  
FT DOMAIN 113 119 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 120 141 S3 OF REPEAT IV (POTENTIAL).  
FT DOMAIN 142 151 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 152 171 S4 OF REPEAT IV (POTENTIAL).  
FT DOMAIN 172 190 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 191 210 S5 OF REPEAT IV (POTENTIAL).  
FT DOMAIN 211 277 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 278 >281 S6 OF REPEAT IV (POTENTIAL).  
FT SITE 244 CALCIUM ION SELECTIVITY AND PERMEABILITY  
FT BINDING <1 18 TO DIHYDROPYRIDINES (BY SIMILARITY).  
FT BINDING 258 >281 TO DIHYDROPYRIDINES (BY SIMILARITY).  
FT BINDING 270 >281 TO PHENYLALKYLAMINES (BY SIMILARITY).  
FT NON\_TER 281 281  
SQ SEQUENCE 281 AA; 32615 MW; 6A07CFFC CRC32;  
  
Query Match 100.0%; Score 49; DB 1; Length 281;  
Best Local Similarity 36.4%; Pred. No. 4.50e+01;  
Matches 4; Conservative 0; Mismatches 7; Indels 0; Gaps 0;  
  
Db 43 PKNPYQYKFWY 53  
QY 2 PXXXXXXFWY 12  
|||  
  
RESULT 7  
ID YCXL\_ODOSI STANDARD; PRT; 355 AA.  
AC P49827;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 01-OCT-1996 (Rel. 34, Last annotation update)  
DE HYPOTHETICAL 41.6 KD PROTEIN IN PSNA-YCF32.1 AND YCF32.2-ACPP  
DE INTERGENIC REGIONS (ORF355).  
OS Odontella sinensis.  
OG Chloroplast.  
OC Eukaryota; stramenopiles; Bacillariophyta; Coscinodiscophyceae;  
OC Bidulphiophycidae; Eupodiscales; Eupodisaceae; Odontella.  
RN SEQUENCE FROM N.A.  
RA KOWALLIK K.V., STOEBE B., SCHAFFRAN I., KROTH-PANIC P., FREIER U.;  
RT "The chloroplast genome of a chlorophyll a+c-containing alga,  
RT Odontella sinensis";  
RL Plant Mol. Biol. Rep. 13:336-342(1995).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration



DE T-CELL RECEPTOR BETA CHAIN V REGION YT35 PRECURSOR.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 84142269.  
RA YANAGI Y., YOSHIKAI Y., LEGGETT K., CLARK S.P., ALEKSANDER I.,  
RA MAK T.W.;  
RT "A human T cell-specific cDNA clone encodes a protein having  
RT extensive homology to immunoglobulin chains.";  
RL Nature 308:145-149(1984).  
DR FIR; A02000; RWHUVY.  
DR HSP; P01789; ZMCP.  
DR PFAM; PF00047; ig; 1.  
KW T-cell; Receptor; Signal.  
FT SIGNAL 1 ?  
FT CHAIN 1 135 T-CELL RECEPTOR BETA CHAIN V REGION YT35.  
FT NON\_TER 135 135  
SQ SEQUENCE 135 AA; 15097 MW; 4080B85f CRC32;  
  
Query Match 100.0%; Score 49; DB 1; Length 135;  
Best Local Similarity 36.4%; Pred. No. 4.50e+01;  
Matches 4; Conservative 0; Mismatches 7; Indels 0; Gaps 0;  
  
Db 44 PISCHNSLEWY 54  
|  
|  
|  
QY 2 PXXXXXXXFWY 12  
  
RESULT 3  
ID TVBL\_MOUSE STANDARD; PRT; 135 AA.  
AC P01734;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 21-JUL-1986 (Rel. 01, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE T-CELL RECEPTOR BETA CHAIN V REGION 3H.25 PRECURSOR.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 85176939.  
RA GOVERNAN J., MINARD K., SHASTRI N., HUNKAPILLER T., HANSBURG D.,  
RA SERCARZ E., HOOD L.;  
RT "Rearranged beta T cell receptor genes in a helper T cell clone  
RT specific for lysozyme: no correlation between V beta and MHC  
RT restriction.";  
RL Cell 40:859-867(1985).  
CC -!- MISCELLANEOUS: THIS T-CELL CLONE EXPRESSES ONLY A SINGLE V-BETA  
CC CHAIN SEGMENT ALTHOUGH IT HAS THREE REARRANGEMENTS IN THE BETA  
CC CHAIN FAMILY.  
CC -!- MISCELLANEOUS: THIS REARRANGED V-BETA CHAIN SEGMENT, SPECIFIC FOR  
CC CHICKEN EGG-WHITE LYSOZYME AND I-A(B), IS THE SAME AS THAT  
CC EXPRESSED IN A HELPER CELL SPECIFIC FOR CYTOCHROME C AND AN  
CC I-E(K) MHC MOLECULE.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
CC EMBL; M12415; AAA40249.1; -  
DR FIR; A02003; RWSBV.  
DR PFAM; PF00047; ig; 1.  
KW T-cell; Receptor; Signal.  
FT SIGNAL 1 20  
FT CHAIN 21 135 T-CELL RECEPTOR BETA CHAIN V REGION  
FT DOMAIN 21 115 3H.25  
V SEGMENT.

FT DOMAIN 116 118 D SEGMENT.  
FT DOMAIN 119 135 J SEGMENT.  
FT DISULFID 42 111 BY SIMILARITY.  
FT NON\_TER 135 135  
SQ SEQUENCE 135 AA; 15123 MW; 03370B4B CRC32;  
  
Query Match 100.0%; Score 49; DB 1; Length 135;  
Best Local Similarity 36.4%; Pred. No. 4.50e+01;  
Matches 4; Conservative 0; Mismatches 7; Indels 0; Gaps 0;  
  
Db 44 PEKHPVFWY 54  
|  
|  
|  
QY 2 PXXXXXXXFWY 12  
  
RESULT 4  
ID YYAR\_BACSU STANDARD; PRT; 173 AA.  
AC P37506;  
DT 01-OCT-1994 (Rel. 30, Created)  
DT 01-OCT-1994 (Rel. 30, Last sequence update)  
DT 01-FEB-1995 (Rel. 31, Last annotation update)  
DE HYPOTHETICAL 20.4 KD PROTEIN IN COTF-TETB INTERGENIC REGION.  
GN YYAR.  
OS Bacillus subtilis.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
OC Bacillus/Staphylococcus group; Bacillus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN-168;  
RX MEDLINE; 96051385.  
RA OGASAWARA N., NAKAI S., YOSHIKAWA H.;  
RT "Systematic sequencing of the 180 kilobase region of the Bacillus  
RT subtilis chromosome containing the replication origin.";  
RL DNA Res. 1:1-14(1994).  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
CC EMBL; D26185; BAA05205.1; -  
DR EMBL; Z99124; CAB16111.1; -  
DR SUBTILIST; BG10033; YYAR.  
DR PFAM; PF00583; Acetyltransf; 1.  
KW Hypothetical protein.  
SQ SEQUENCE 173 AA; 20393 MW; E0EDA2A6 CRC32;  
  
Query Match 100.0%; Score 49; DB 1; Length 173;  
Best Local Similarity 36.4%; Pred. No. 4.50e+01;  
Matches 4; Conservative 0; Mismatches 7; Indels 0; Gaps 0;  
  
Db 160 PTANEIAFWY 170  
|  
|  
|  
QY 2 PXXXXXXXFWY 12  
  
RESULT 5  
ID T2ML\_MYCSP STANDARD; PRT; 202 AA.  
AC P43642;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 01-NOV-1995 (Rel. 32, Last annotation update)  
DE TYPE II RESTRICTION ENZYME MUNI (EC 3.1.21.4) (ENDONUCLEASE MUNI)  
DE (R.MUNI).  
GN MUNIR.  
OS Mycoplasma sp.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;  
OC Mycoplasmataceae; Mycoplasma.  
RN [1]  
RP SEQUENCE FROM N.A.

\*\*\*\*\*  
M P S R C H  
\*\*\*\*\* (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Sat Apr 15 01:42:23 2000; MasPar time 3.05 Seconds  
Tabular output not generated. 117.497 Million cell updates/sec

Title: >US-08-452-843-25  
Description: (1-12) from US08452843.pep  
Perfect score: 49  
Sequence: 1 XPXXXXXXFWY 12

Scoring table: PAM 150  
Gap 15

Searched: 82229 seqs, 29864866 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: swiss-prot38  
1:swissprot

Statistics: Mean 22.239; Variance 40.519; scale 0.549

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES									
Result No.	Query			DB	ID	Description	Pred. No.		
	Score	Match	Length						
1	49	100.0	134	1	TVB7_MOUSE	T-CELL RECEPTOR BETA C	4.50e+01		
2	49	100.0	135	1	TVB1_HUMAN	T-CELL RECEPTOR BETA C	4.50e+01		
3	49	100.0	135	1	TVB1_MOUSE	T-CELL RECEPTOR BETA C	4.50e+01		
4	49	100.0	173	1	YVAR_BACSU	HYPOTHETICAL 20.4 KD P	4.50e+01		
5	49	100.0	202	1	T2MU_MYCSP	TYPE II RESTRICTION EN	4.50e+01		
6	49	100.0	281	1	CCAD_MOUSE	VOLTAGE-DEPENDENT L-TY	4.50e+01		
7	49	100.0	355	1	VCX1_ODOST	HYPOTHETICAL 41.6 KD P	4.50e+01		
8	49	100.0	368	1	GALT_HUMAN	GALANIN RECEPTOR TYPE	4.50e+01		
9	49	100.0	370	1	GALT_MOUSE	GALANIN RECEPTOR TYPE	4.50e+01		
10	49	100.0	370	1	GALT_RAT	GALANIN RECEPTOR TYPE	4.50e+01		
11	49	100.0	396	1	MALE_ECOLI	MALTOSE-BINDING PERIPL	4.50e+01		
12	49	100.0	396	1	MALE_ENTAE	MALTOSE-BINDING PERIPL	4.50e+01		
13	49	100.0	396	1	MALE_SALTY	MALTOSE-BINDING PERIPL	4.50e+01		
14	49	100.0	436	1	PPAL_ASPNG	PHOSPHATE-REPRESSIBLE	4.50e+01		
15	49	100.0	452	1	PCNB_HAEIN	PROBABLE POLY(A) POLYM	4.50e+01		
16	49	100.0	472	1	PCNB_ECOLI	POLY(A) POLYMERASE (EC	4.50e+01		
17	49	100.0	590	1	PHO4_NEUCR	PHOSPHATE-REPRESSIBLE	4.50e+01		
18	49	100.0	688	1	CACM_YEAST	PUTATIVE MITOCHONDRIAL	4.50e+01		
19	49	100.0	728	1	GIGCB_ECOLI	1,4-ALPHA-GLUCAN BRANC	4.50e+01		
20	49	100.0	796	1	YH04_YEAST	HYPOTHETICAL 91.2 KD P	4.50e+01		
21	49	100.0	867	1	POL_IPMA	PUTATIVE POLY(POLYPROTE	4.50e+01		
22	49	100.0	1517	1	YD22_SCHPO	HYPOTHETICAL 170.7 KD	4.50e+01		
23	49	100.0	1610	1	CCAD_MESAU	VOLTAGE-DEPENDENT L-TY	4.50e+01		

24	49	100.0	1815	1	UN13_CABEL	1	PHORBOL ESTER/DIACYGL	4.50e+01
25	49	100.0	2161	1	CCAD_HUMAN	1	VOLTAGE-DEPENDENT L-TY	4.50e+01
26	49	100.0	2190	1	CCAD_CHICK	1	VOLTAGE-DEPENDENT L-TY	4.50e+01
27	49	100.0	2203	1	CCAD_RAT	1	VOLTAGE-DEPENDENT L-TY	4.50e+01
28	49	100.0	3176	1	CA36_HUMAN	1	COLLAGEN ALPHA 3(VI) C	4.50e+01
29	45	91.8	125	1	YCFL_ECOLI	1	HYPOTHETICAL 14.0 KD P	1.64e+02
30	45	91.8	133	1	TVB2_HUMAN	1	T-CELL RECEPTOR BETA C	1.64e+02
31	45	91.8	184	1	YBET_ECOLI	1	HYPOTHETICAL 20.9 KD P	1.64e+02
32	45	91.8	302	1	FSTL_ARATH	1	FLAVONOL SULFOTRANSFER	1.64e+02
33	45	91.8	321	1	TCB_FLY	1	T-CELL RECEPTOR BETA C	1.64e+02
34	45	91.8	332	1	HTR1_HAEIN	1	IRON-UTILIZATION PERIP	1.64e+02
35	45	91.8	365	1	VG13_BPPH2	1	MORPHOGENESIS PROTEIN	1.64e+02
36	45	91.8	402	1	VGLD_PVRV1	1	GLYCOPROTEIN GP50	1.64e+02
37	45	91.8	424	1	I131_MOUSE	1	INTERLEUKIN-13 RECEPTO	1.64e+02
38	45	91.8	451	1	PHT1_PSEPU	1	PTHALATE TRANSPORTER	1.64e+02
39	45	91.8	572	1	LAC3_THACU	1	LACCASE 3 PRECURSOR (E	1.64e+02
40	45	91.8	576	1	LAC1_THACU	1	LACCASE 1 PRECURSOR (E	1.64e+02
41	45	91.8	599	1	LAC2_THACU	1	LACCASE 2 PRECURSOR (E	1.64e+02
42	45	91.8	608	1	YD56_YEAST	1	PUTATIVE MULTICOPPER O	1.64e+02
43	45	91.8	881	1	YFCU_ECOLI	1	HYPOTHETICAL OUTER MEM	1.64e+02
44	45	91.8	3106	1	LMA2_MOUSE	1	LAMININ ALPHA-2 CHAIN	1.64e+02
45	45	91.8	3110	1	LMA2_HUMAN	1	LAMININ ALPHA-2 CHAIN	1.64e+02

ALIGNMENTS

RESULT	1	STANDARD;	PRT;	134 AA.
ID	TVB7_MOUSE			
AC	P06320;			
DT	01-JAN-1988 (Rel. 06, Created)			
DT	01-JAN-1988 (Rel. 06, Last sequence update)			
DT	15-JUL-1999 (Rel. 38, Last annotation update)			
DE	T-CELL RECEPTOR BETA CHAIN V REGION CTL-F3 PRECURSOR.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;			
OC	Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE; 87053852.			
RA	CHOU H.S., BEHLKE M.A., GODAMBE S.A., RUSSELL J.H., BROOKS C.G.,			
RA	LOH D.Y.;			
RT	"T cell receptor genes in an alloreactive CTL clone: implications for			
RT	rearrangement and germline diversity of variable gene segments.";			
RL	EMBO J. 5:2149-2155(1986).			
DR	PIR; A02002; RWSB3.			
DR	PFAM; PF00047; ig: 1.			
KW	T-cell; Receptor; Glycoprotein; Signal.			
FT	SIGNAL	1	19	
FT	CHAIN	20	134	
FT	DOMAIN	20	115	
FT	DOMAIN	116	119	
FT	DOMAIN	120	134	
FT	DISULFID	42	111	
FT	CARBOHYD	90	90	
FT	NON_TER	134	134	
SQ	SEQUENCE	134 AA;	14946 MW;	05686B71 CRC32;

Query Match 100.0%; Score 49; DB 1; Length 134;  
Best Local Similarity 36.4%; Pred. No. 4.50e+01;  
Matches 4; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Db 44 PISGHSVAFWY 54  
QY 2 PXXXXXXFWY 12

RESULT 2  
ID TVB1\_HUMAN STANDARD; PRT; 135 AA.  
AC P01733;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 21-JUL-1986 (Rel. 01, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)

Db 78 PNTKSCARFWY 88  
QY 2 PXXXXXXFWY 12

RESULT 15

ID R65451 standard; protein; 113 AA.  
AC R65451;  
DE 24-MAY-1995 (first entry)  
DE T-cell receptor V-beta HVB6.2/3.  
KW T-cell receptor; TCR; T-lymphocyte receptor; variable region;  
KW beta-chain; V-beta; multiple sclerosis; cerebrospinal fluid;  
KW autoimmune disease; lymphoma; vaccine.  
OS Homo sapiens.  
PN WO9425063-A.  
PD 10-NOV-1994.  
PF 29-APR-1994; U04789.  
PR 29-APR-1993; US-055006.  
PA (IMMU-) IMMUNE RESPONSE CORP.  
PA (SAND-) SAN DIEGO REGIONAL CANCER CENT.  
PI Brostoff SW, Carlo DJ, Gold DP, Smith LR, Wilson DB;  
DR WPI; 94-357913/44.  
PT New vaccine against multiple sclerosis using T-cell receptors  
PT or fragments of T-cell receptors from the beta chain variable  
PT region; for treating autoimmune disease and lymphoma(s)  
PS Disclosure; Fig. 2A; 43pp; English.  
CC Sequences of the T-cell receptor beta-chain variable region that  
CC were most frequently expressed in cultures from the cerebrospinal  
CC fluid of multiple sclerosis patients are given in R65450-67. A  
CC peptide based on R65450-57 has been used for vaccine development.  
SQ Sequence 113 AA;

Query Match 100.0%; Score 49; DB 1; Length 113;  
Best Local Similarity 36.4%; Pred. No. 3.66e+02;  
Matches 4; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Db 44 PISGHVSLFWY 54  
QY 2 PXXXXXXFWY 12

Search completed: Sat Apr 15 01:41:30 2000  
Job time : 39 secs.

(FARB ) BAYER AG.  
PI Chu M-L, Ebberts J, Hoerlein D, Timpl R;  
DR WPI: 94-025477/03.  
DR N-PSDB; Q53999.  
PT New Kunitz-type proteinase inhibitor - having aminoacid changes  
PT in inhibitor derived from alpha 3-chain of human type VI  
PT collagen, used to treat emphysema, coagulation disorders etc.  
PS Example; Fig 2; 9pp; English.  
CC The inventors claim a new Kunitz type proteinase inhibitor whose AA  
CC sequence is deduced from cDNA clones which cover 3kb of the type VI  
CC collagen alpha-3 chain mRNA. The cDNA was isolated from a placenta  
CC and a fibroblast cDNA library. The selected proteinase inhibitor  
CC domain is located in the segment given in R47542. Q53999 encodes  
CC the alpha-3(VI)-Kunitz type inhibitor modified for gene expression  
CC by 5' and 3' extensions. The extensions contain HindIII and  
CC BamHI restriction sites and the sequence encoding the KEX2  
CC processing site of the alpha-mating factor leader sequence. The  
CC KEX2-enzyme processing site is marked on FT R47543.  
SQ Sequence 63 AA;  
  
Query Match 100.0%; Score 49; DB 1; Length 63;  
Best Local Similarity 36.4%; Pred. No. 3.66e+02;  
Matches 4; Conservative 0; Mismatches 7; Indels 0; Gaps 0;  
  
Db 30 PNTKSCARFWY 40  
|  
|  
|  
QY 2 PXXXXXXXFWY 12  
  
RESULT 12  
ID R47542 standard; Protein; 70 AA.  
AC R47542;  
DT 21-JUL-1994 (first entry)  
DE Sequence of the last 70 AAs of the C5 domain of the type VI  
DE collagen alpha-3-chain (AAs 2873-2943) contg. the Kunitz type  
DE inhibitor domain.  
KW Kunitz type inhibitor domain; type VI collagen alpha-3 chain;  
KW serin proteinase inhibitor.  
OS Homo sapiens.  
PN US278285-A.  
PD 11-JAN-1994.  
PF 01-FEB-1990; 473295.  
PR 01-FEB-1990; US-473295.  
PA (FARB ) BAYER AG.  
PI Chu M-L, Ebberts J, Hoerlein D, Timpl R;  
DR WPI: 94-025477/03.  
DR N-PSDB; Q53998.  
PT New Kunitz-type proteinase inhibitor - having aminoacid changes  
PT in inhibitor derived from alpha 3-chain of human type VI  
PT collagen, used to treat emphysema, coagulation disorders etc.  
PS Disclosure; Fig 1; 9pp; English.  
CC The inventors claim a new Kunitz type proteinase inhibitor whose AA  
CC sequence is deduced from cDNA clones which cover 3kb of the type VI  
CC collagen alpha-3 chain mRNA. The cDNA was isolated from a placenta  
CC and a fibroblast cDNA library. The selected proteinase inhibitor  
CC domain is located in the segment given in R47542. A modified Kunitz-  
CC type proteinase inhibitor which is claimed consists of the AA  
CC sequence in R47542 modified at posns. 16, 17 and 39 as follows:  
CC Ala16-Arg17-Arg39.  
SQ Sequence 70 AA;  
  
Query Match 100.0%; Score 49; DB 1; Length 70;  
Best Local Similarity 36.4%; Pred. No. 3.66e+02;  
Matches 4; Conservative 0; Mismatches 7; Indels 0; Gaps 0;  
  
Db 25 PNTKSCARFWY 35  
|  
|  
|  
QY 2 PXXXXXXXFWY 12  
  
RESULT 13  
ID W73960 standard; Protein; 100 AA.  
AC W73960;

DT 29-APR-1999 (first entry)  
DE Human TPC2/MBP fusion protein.  
KW TPC2; TPC3; human; telomere length regulation; cancer; pregnancy; MBP;  
KW fertility; diagnosis; therapy; fusion protein; maltose binding protein.  
OS Synthetic.  
OS Homo sapiens.  
PN US585877-A.  
PD 12-JAN-1999.  
PF 13-SEP-1996; 710249.  
PR 08-SEP-1995; US-003492.  
PR 05-JAN-1996; US-563808.  
PR 13-SEP-1996; US-710249.  
PA (GERO-) GERON CORP.  
PI Adams RR, Andrews WH, Feng J, Villeponteau B;  
DR WPI: 99-152104/13.  
PT DNA encoding proteins TPC2 and TPC3 - useful for regulating telomere  
PT length or modulating telomerase activity  
PS Disclosure; Column 27; 59pp; English.  
CC This sequence represents a fusion protein between the human TPC2 protein  
CC and the human maltose binding protein (MBP), which can be contained  
CC within the recombinant mammalian host cell of the invention. The  
CC invention provides methods and reagents for regulating telomere length  
CC and modulating telomerase activity in mammalian cells as well as for  
CC detecting, diagnosing, and treating related diseases and conditions such  
CC as cancer, pregnancy, or fertility in humans and other mammals.  
SQ Sequence 100 AA;  
  
Query Match 100.0%; Score 49; DB 1; Length 100;  
Best Local Similarity 36.4%; Pred. No. 3.66e+02;  
Matches 4; Conservative 0; Mismatches 7; Indels 0; Gaps 0;  
  
Db 1 PNIPQMSAFWY 11  
|  
|  
|  
QY 2 PXXXXXXXFWY 12  
  
RESULT 14  
ID R39681 standard; Protein; 111 AA.  
AC R39681;  
DT 13-JAN-1994 (first entry)  
DE Protein encoded by pKFN-1745 412 bp EcoRI-XbaI fragment.  
KW Human type VI collagen; preferential inhibition; cathepsin G;  
KW neutrophil elastase; protease 3; trypsin; plasmin; kallikrein;  
KW pathological proteolysis treatment; polymorphonuclear leucocytes;  
KW enzyme release; acute pancreatitis; inflammation; thrombocytopaenia;  
KW rheumatoid arthritis; psoriasis; organ preservation;  
KW platelet function preservation; domain; plasmid pKFN-1000.  
OS Synthetic.  
FH Key  
FT peptide 1. .53 Location/Qualifiers  
FT peptide /note= "signal peptide"  
FT peptide 54..111  
FT peptide /note= "mature peptide"  
FT WO9314119-A.  
PN 22-JUL-1993.  
PD 07-JAN-1993; DK0002.  
PR 07-JAN-1992; WO-DK0005.  
PA (NOVO) NOVO-NORDISK AS.  
PI Bjorn SE, Norris F, Norris K, Olsen OH, Petersen LC;  
DR WPI: 93-243147/30.  
DR N-PSDB; Q46652.  
PT Human Kunitz-type protease inhibitor variants - for treating and  
PT preventing diseases associated with pathological proteolysis e.g.  
PT acute pancreatitis, inflammation, thrombocytopaenia etc.  
PS Example; Page 24; 39pp; English.  
CC The sequence is that encoded by the 412 bp EcoRI-XbaI fragment  
CC from pKFN-1745 which was used in the prodn. of human alpha3  
CC (VI) Kunitz-type protease inhibitor domain (see R39669).  
SQ Sequence 111 AA;  
  
Query Match 100.0%; Score 49; DB 1; Length 111;  
Best Local Similarity 36.4%; Pred. No. 3.66e+02;  
Matches 4; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

DE Human aprotinin-like Kunitz domain (A3 collagen).  
 KW Aprotinin; Kunitz domain; human neutrophil elastase; hNE;  
 KW connective tissue; alpha 1 protease inhibitor; API; neutrophil;  
 KW alpha 1 antitrypsin; respiratory disorder; cystic fibrosis;  
 OS smokers emphysema.  
 OS Homo sapiens.  
 PN WO9620278-A2.  
 PD 04-JUL-1996.  
 PF 15-DEC-1995; U16349.  
 PR 16-DEC-1994; US-358160.  
 PA (PROT-) PROTEIN ENG CORP.  
 PI Guterman SK, Kent RB, Ladner RC, Ley AC, Markland W;  
 PI Roberts BL.  
 DR WPI; 96-321851/32.  
 PT New engineered inhibitors of human neutrophil elastase - contg.  
 PT aprotinin-like Kunitz domain for treating, e.g. cystic fibrosis or  
 PT other respiratory disorders  
 PS Disclosure; Page 53; 105pp; English.  
 CC Genetically engineered human derived Kunitz domains can be used to  
 CC inhibit human neutrophil elastase, an enzyme involved in the  
 CC elimination of pathogens and the restructuring of connective tissue.  
 CC In cases of reduction of the circulating alpha-1-protease inhibitor  
 CC (API or alpha 1 antitrypsin), or the inactivation of API by oxidation  
 CC (smokers emphysema), extensive destruction of the lung tissue may  
 CC result from uncontrolled elastolytic activity of human neutrophil  
 CC elastase. Other respiratory disorders such as cystic fibrosis are  
 CC thought to be caused by human neutrophil elastase release by  
 CC neutrophils. The genetically engineered human derived Kunitz  
 CC domains can be used to treat such respiratory disorders. See  
 CC R99146-R99211.  
 SQ Sequence 58 AA;

Query Match 100.0%; Score 49; DB 1; Length 58;  
 Best Local Similarity 36.4%; Pred. No. 3.66e+02;  
 Matches 4; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Db 25 PNTKSCARFWY 35  
 |  
 QY 2 PXXXXXXFWY 12

RESULT 9  
 ID R39671 standard; protein; 58 AA.  
 AC R39671;  
 DT 13-JAN-1994 (first entry)  
 DE C-terminal Kunitz-type protease inhibitor variant.  
 KW Human type VI collagen; preferential inhibition; cathepsin G;  
 KW neutrophil elastase; protease 3; trypsin; plasmin; kallikrein;  
 KW pathological proteolysis treatment; polymorphonuclear leucocytes;  
 KW enzyme release; acute pancreatitis; inflammation; thrombocytopaenia;  
 KW rheumatoid arthritis; psoriasis; organ preservation;  
 KW platelet function preservation; domain; example.  
 OS Homo sapiens.  
 PN WO9314119-A.  
 PD 22-JUL-1993.  
 PF 07-JAN-1993; DR0002.  
 PR 07-JAN-1992; WO-DK0005.  
 PA (NOVO ) NOVO-NORDISK AS.  
 PI Bjorn SE, Norris F, Norris K, Olsen OH, Petersen LC;  
 DR WPI; 93-243147/30.  
 PT Human Kunitz-type protease inhibitor variants - for treating and  
 PT preventing diseases associated with pathological proteolysis e.g.  
 PT acute pancreatitis, inflammation, thrombocytopaenia etc.  
 PS Example; Page 26; 33pp; English.  
 CC The sequence is that of a variant of the C-terminal Kunitz-type  
 CC protease inhibitor domain of the alpha3 chain of human type VI  
 CC collagen. Modification of the Kunitz sequence provides preferential  
 CC inhibition of e.g. neutrophil elastase, cathepsin G, protease-3,  
 CC trypsin, plasmin and/or kallikrein. The variant can thus be used to  
 CC treat pathological proteolysis caused by enzymes released from  
 CC polymorphonuclear leucocytes, e.g. acute pancreatitis, inflammation,  
 CC thrombocytopaenia, rheumatoid arthritis, psoriasis, etc., opt.  
 CC administered with heparin. It can also be used for preservation of

CC organs or platelet function, and possibly for isolating natural cpds.  
 CC which bind to the specified domain. It has zero net charge and so  
 CC is less likely to cause kidney damage than aprotinin and, being of  
 CC human origin is less immunogenic.  
 SQ Sequence 58 AA;

Query Match 100.0%; Score 49; DB 1; Length 58;  
 Best Local Similarity 36.4%; Pred. No. 3.66e+02;  
 Matches 4; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Db 25 PNTKSCARFWY 35  
 |  
 QY 2 PXXXXXXFWY 12

RESULT 10  
 ID W64117 standard; peptide; 58 AA.  
 AC W64117;  
 DT 06-OCT-1998 (first entry)  
 DE Human Kunitz-type serine protease inhibitor domain #13.  
 KW Kunitz domain; serine protease inhibitor; kallikrein; plasma; treatment;  
 KW inflammation; septic shock; hypotension; post-operative bleeding;  
 KW adult respiratory distress syndrome; Factor Xla;  
 KW disseminated intravascular coagulation.  
 OS Homo sapiens.  
 PN US5786328-A.  
 PD 28-JUN-1998.  
 PF 05-JUN-1995; 463432.  
 PR 05-JUN-1995; US-463432.  
 PA (GETH ) GENENTECH INC.  
 PI Dennis MS, Lazarus RA;  
 DR WPI; 98-436581/37.  
 PT Inhibition of plasma kallikrein in vivo - using polypeptide  
 PT comprising non-native Kunitz-type serine protease inhibitor domain  
 PS Disclosure; Column 47-48; 41pp; English.  
 CC W64078-W64149 are Kunitz-type serine protease inhibitor domains. These  
 CC protein fragments could be used for treating a mammal for which  
 CC inhibition of plasma kallikrein is indicated. The peptides can be used  
 CC for treating inflammation, septic shock, hypotension, adult respiratory  
 CC distress syndrome, disseminated intravascular coagulation and  
 CC postoperative bleeding. Modified peptides have increased potency for  
 CC plasma kallikrein and also inhibit Factor Xla serine protease.  
 SQ Sequence 58 AA;

Query Match 100.0%; Score 49; DB 1; Length 58;  
 Best Local Similarity 36.4%; Pred. No. 3.66e+02;  
 Matches 4; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Db 25 PNTKSCARFWY 35  
 |  
 QY 2 PXXXXXXFWY 12

RESULT 11  
 ID R47543 standard; Protein; 63 AA.  
 AC R47543;  
 DT 21-JUL-1994 (first entry)  
 DE Sequence encoded by alpha-3(VI)-Kunitz type inhibitor cDNA modified  
 DE for gene expression with additional 5' and 3' extensions.  
 KW Kunitz type inhibitor; type VI collagen alpha-3 chain;  
 KW serine proteinase inhibitor.  
 OS Synthetic.  
 FH Key  
 FT peptide  
 FT /label= alpha-F-leader  
 FT 6 63  
 FT /label= alpha-3(VI)inhibitor  
 FT 5 6  
 FT /label= KEX2 processing site  
 PN US5278285-A.  
 PD 11-JAN-1994.  
 PF 01-FEB-1990; 473295.  
 PR 01-FEB-1990; US-473295.

CC and/or plasmin to Tissue Factor (TF) may provide an effective therapy.  
CC Serine protease inhibitor polypeptides reversibly inhibit the formation  
CC of TF-Factor VII complexes (also complexes involving TF and Factor Xa,  
CC plasma kallikrein and/or plasmin) by binding to the active site of the  
CC TF, and therefore preventing it reacting with the serine proteases. The  
CC formation of a TF-Factor VII complex is the initiating step in the  
CC cascade of reactions that leads to blood coagulation.  
SQ Sequence 58 AA;

Query Match 100.0%; Score 49; DB 1; Length 58;  
Best Local Similarity 36.4%; Pred. No. 3.66e+02;  
Matches 4; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Db 25 PNTKSCARFWY 35  
|  
|  
|  
QY 2 PXXXXXXXFWY 12

## RESULT 5

ID R39672 standard; protein; 58 AA.  
AC R39672;  
DT 13-JAN-1994 (first entry)  
DE C-terminal Kunitz-type protease inhibitor variant.  
KW Human type VI collagen; preferential inhibition; cathepsin G;  
KW neutrophil elastase; protease 3; trypsin; plasmin; kallikrein;  
KW pathological proteolysis treatment; polymorphonuclear leucocytes;  
KW enzyme release; acute pancreatitis; inflammation; thrombocytopaenia;  
KW rheumatoid arthritis; psoriasis; organ preservation;  
KW platelet function preservation; domain; example.  
OS Homo sapiens.  
PN WO9314119-A.  
PD 22-JUL-1993.  
PF 07-JAN-1993; DK0002.  
PR 07-JAN-1992; WO-DK0005.  
PA (NOVO) NOVO-NORDISK AS.  
PI Bjorn SE, Norris F, Norris K, Olsen OH, Petersen LC;  
DR WPI; 93-243147/30.  
PT Human Kunitz-type protease inhibitor variants - for treating and  
PT preventing diseases associated with pathological proteolysis e.g.  
PT acute pancreatitis; inflammation; thrombocytopaenia etc.  
PS Example; Page 26; 33pp; English.  
CC The sequence is that of a variant of the C-terminal Kunitz-type  
CC protease inhibitor domain of the alpha3 chain of human type VI  
CC collagen. Modification of the Kunitz sequence provides preferential  
CC inhibition of e.g. neutrophil elastase, cathepsin G, protease-3,  
CC trypsin, plasmin and/or kallikrein. The variant can thus be used to  
CC treat pathological proteolysis caused by enzymes released from  
CC polymorphonuclear leucocytes, e.g. acute pancreatitis, inflammation,  
CC thrombocytopaenia, rheumatoid arthritis, psoriasis, etc., Opt.  
CC administered with heparin. It can also be used for preservation of  
CC organs or platelet function, and possibly for isolating natural cpds.  
CC which bind to the specified domain. It has zero net charge and so  
CC is less likely to cause kidney damage than aprotinin and, being of  
CC human origin is less immunogenic.  
SQ Sequence 58 AA;

Query Match 100.0%; Score 49; DB 1; Length 58;  
Best Local Similarity 36.4%; Pred. No. 3.66e+02;  
Matches 4; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Db 25 PNTKSCARFWY 35  
|  
|  
|  
QY 2 PXXXXXXXFWY 12

## RESULT 6

ID R81918 standard; protein; 58 AA.  
AC R81918;  
DT 18-MAR-1996 (first entry)  
DE Human collagen alpha-3 Kunitz domain.  
KW Human collagen alpha-3; kallikrein;  
KW inhibitor; KIP; Kunitz domain; hereditary angioedema.  
OS Homo sapiens.

PN WO9521601-A2.  
PD 17-AUG-1995.  
PF 11-JAN-1995; U00299.  
PR 11-JAN-1994; US-179964.  
PR 10-MAR-1994; US-208264.  
PA (PROT-) PROTEIN ENG CORP.  
PI Ladner RC, Markland W;  
DR WPI; 95-292934/38.  
PT Kallikrein inhibiting proteins comprising a Kunitz domain homologous  
PT to bovine pancreatic trypsin inhibitor - useful for preventing or  
PT treating disorders attributable to excessive kallikrein activity,  
PT eg. in hereditary angioedema.  
PS Disclosure; Page 27; 46pp; English.  
CC R81918 is the human collagen alpha-3 Kunitz domain, a kallikrein  
CC inhibiting protein (KIP). The KIP can be used for treating or  
CC preventing disorders attributable to excessive kallikrein activity,  
CC e.g. hereditary angioedema. The KIP can also be used for assaying,  
CC purifying and in vivo imaging of kallikrein.  
SQ Sequence 58 AA;

Query Match 100.0%; Score 49; DB 1; Length 58;  
Best Local Similarity 36.4%; Pred. No. 3.66e+02;  
Matches 4; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Db 25 PNTKSCARFWY 35  
|  
|  
|  
QY 2 PXXXXXXXFWY 12

## RESULT 7

ID R78545 standard; peptide; 58 AA.  
AC R78545;  
DT 01-MAR-1996 (first entry)  
DE Human collagen alpha-3 Kunitz domain.  
KW Human; lipoprotein-associated coagulation inhibitor; peptide library;  
KW inhibitor; plasmin; bovine; pancreatic trypsin inhibitor; Kunitz domain;  
KW fibrinolysis; fibrinogenolysis; bleeding; thrombolytic.  
OS Homo sapiens.  
PN WO9518830-A2.  
PD 13-JUL-1995.  
PF 11-JAN-1995; U00298.  
PR 11-JAN-1994; US-179658.  
PR 10-MAR-1994; US-208265.  
PA (PROT-) PROTEIN ENG CORP.  
PI Ladner RC, Markland W;  
DR WPI; 95-255042/33.  
PT Novel plasmin inhibiting protein comprising a Kunitz Domain - useful  
PT to prevent/treat disorders attributable to excess plasmin activity.  
PS Claim 3; Page 34; 59pp; English.  
CC The peptides R78435-R78570 are derivatives of the Kunitz domains from a  
CC variety of plasmin inhibitors e.g. the human lipoprotein-associated  
CC coagulation inhibitor (LACI) Kunitz domains 1, 2 or 3. The peptides were  
CC designed based on the Kunitz domains and are named Designed Plasmin  
CC inhibitor (DPI). This peptide is the human collagen alpha-3 Kunitz  
CC domain (KUDOM).

CC The peptides can be used to prevent or treat a clinical condition  
CC exacerbated by plasmin e.g. inappropriate fibrinolysis or  
CC fibrinogenolysis, excessive bleeding associated with thrombolytics.  
SQ Sequence 58 AA;

Query Match 100.0%; Score 49; DB 1; Length 58;  
Best Local Similarity 36.4%; Pred. No. 3.66e+02;  
Matches 4; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Db 25 PNTKSCARFWY 35  
|  
|  
|  
QY 2 PXXXXXXXFWY 12

## RESULT 8

ID R99204 standard; protein; 58 AA.  
AC R99204;  
DT 12-FEB-1997 (first entry)

```

RESULT 2
ID R39675 standard; protein: 56 AA.
AC R39675;
DT 13-JAN-1994 (first entry)
DE C-terminal Kunitz-type protease inhibitor variant.
KW Human type VI collagen; preferential inhibition; cathepsin G;
KW neutrophil elastase; protease 3; trypsin; plasmin; kallikrein;
KW pathological proteolysis treatment; polymorphonuclear leucocytes;
KW enzyme release; acute pancreatitis; inflammation; thrombocytopenia;
KW rheumatoid arthritis; psoriasis; organ preservation;
KW platelet function preservation; domain; example.
OS Homo sapiens.
PN W09314119-A.
PD 22-JUL-1993.
PF 07-JAN-1993; DK0002.
PR 07-JAN-1992; WO-DK0005.
PA (NOVO) NOVO-NORDISK AS.
PI Bjorn SE, Norris F, Norris K, Olsen OH, Petersen LC;
DR WPI; 93-243147/30.
PT Human Kunitz-type protease inhibitor variants - for treating and
PT preventing diseases associated with pathological proteolysis e.g.
PT acute pancreatitis, inflammation, thrombocytopenia etc.
PS Example; Page 26; 33pp; English.
CC The sequence is that of a variant of the C-terminal Kunitz-type
CC protease inhibitor domain of the alpha3 chain of human type VI
CC collagen. Modification of the Kunitz sequence provides preferential
CC inhibition of e.g. neutrophil elastase, cathepsin G, protease-3,
CC trypsin, plasmin and/or kallikrein. The variant can thus be used to
CC treat pathological proteolysis caused by enzymes released from
CC polymorphonuclear leucocytes, e.g. acute pancreatitis, inflammation,
CC thrombocytopenia, rheumatoid arthritis, psoriasis, etc., Opt.
CC administered with heparin. It can also be used for preservation of
CC organs or platelet function, and possibly for isolating natural cpds.
CC which bind to the specified domain. It has zero net charge and so
CC is less likely to cause kidney damage than aprotinin and, being of
CC human origin is less immunogenic.
CC Sequence 56 AA;

Query Match 100.0%; Score 49; DB 1; Length 56;
Best Local Similarity 36.4%; Pred. No. 3.66e+02;
Matches 4; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Db 23 PNTKSCARFWY 33
QY 2 PXXXXXXXFWY 12

RESULT 4
ID W92864 standard; peptide: 58 AA.
AC W92864;
DT 17-MAY-1999 (first entry)
DE US5880256 Seq ID 40.
KW Serine protease inhibitor; Factor VIIA inhibitor; Factor XIA inhibitor;
KW plasma kallikrein; plasmin; Kunitz-type; reocclusion; thrombolysis;
KW thrombosis; thrombophlebitis; stroke; infarction; coagulation; clotting;
KW respiratory distress syndrome; clotting; bleeding; surgery; angioplasty;
KW pulmonary embolism; inflammation; arteriosclerosis; transient ischemia;
KW coronary artery disease; angina; arterial fibrillation; blood; thrombi;
KW emboli; TP-Factor VII complex; tissue factor.
OS Homo sapiens.
PN US5880256-A.
PD 09-MAR-1999.
PF 03-MAR-1995; 399115.
PR 03-MAR-1995; US-399115.
PR 04-MAR-1994; US-206310.
PA (GETH) GENENTECH INC.
PI Dennis MS, Lazarus RA;
DR WPI; 99-204039/17.
PT New polypeptides containing Kunitz-type serine protease inhibitor
PT domains - useful for treating hypercoagulation leading to thrombi
PT and emboli
PS Disclosure: Column 49-50; 47pp; English.
CC This invention describes novel polypeptides comprising a Kunitz-type
CC serine protease inhibitor domain, with an equilibrium dissociation
CC constant (K1) of less than 100 nM for tissue factor-Factor VIIa
CC complexes. The peptides have the formula: R1-X11-X12-X13-X14-X15-X16-
CC X17-X18-X19-R2-X34-R3-X38-X39-R4 where R1- a 5-10 amino acids residues
CC (aa) peptide, at least 1 aa is Cys; R2- a 14 aa peptide, at least 1 aa
CC is Cys; R3- a tripeptide; R4- a 12-19 aa peptide, at least 1 aa is Cys;
CC X11- Pro, Arg, Ala, Glu, Gly or Thr; X12- Gly; X13- Pro, Leu, Trp, Val,
CC X14- Phe, His, Tyr, Ala, Ile, Glu or Gin; X15- Cys, Ala, Ser, Thr or Gly;
CC X16- Gly or Ala; X17- Met, Leu, Ile, Arg, Tyr or
CC X18- Ile, His, Leu, Met, Tyr or Phe; X19- Leu, Arg, Ala, Lys or Ile;
CC Ser; X34- Phe, Ile, Ser, Tyr, Trp or Val; X38- Cys, Ala, Ser, Thr or Gly;
CC X39- Tyr, Gly, Trp, His or Phe; provided that: R1 is not X1-Asp-Ile-
CC Cys-Lys-Leu-Pro-Lys-Asp where X1- His or a 1-5 aa peptide (e.g. W92825)
CC and X11-X19 are not: Pro-Gly-Phe-Ala-Lys-Ala-Ile-Ile-Arg (e.g. W92826),
CC Thr-Gly-Leu-Cys-Lys-Ala-Tyr-Ile-Arg (e.g. W92827) Thr-Gly-Leu-Cys-Lys-
CC Ala-Arg-Ile-Arg (e.g. W92828) and Ala-Gly-Ala-Ala-Lys-Ala-Leu-Leu-Ala
CC (e.g. W92829). Such peptides may be used to treat arterial reocclusion
CC after thrombolysis, venous thrombosis, deep venous thrombophlebitis,
CC stroke, infarction, disseminated intravascular coagulation (DIC), adult
CC respiratory distress syndrome (ARDS), clotting, bleeding, surgery,
CC percutaneous transluminal coronary angioplasty, pulmonary embolism,
CC inflammation, arteriosclerosis, coronary artery disease, angina,
CC artificial heart valves, transient ischemia, arterial fibrillation and
CC other conditions caused by over activity of the blood clotting process
CC (hypercoagulation leading to intravascular thrombi and emboli) in which
CC inhibiting the binding of Factor VII, Factor Xla, plasma kallikrein

```

\*\*\*\*\*

MPERCH\_PP

(TM)

\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.

Copyright (c) 1993-1998 University of Edinburgh, U.K.

Distribution rights by Oxford Molecular Ltd

MPPerch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Sat Apr 15 01:40:51 2000; MasPar time 3.13 Seconds

Tabular output not generated. 90.748 Million cell updates/sec

Title: >US-08-452-843-25

Description: (1-12) from US08452843.pep

Perfect Score: 49

Sequence: 1 XPXXXXXXFWY 12

Scoring table: PAM 150

Gap 15

Searched: 188963 seqs, 23686106 residues

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database: a-geneseq36  
1:geneseqp

Statistics: Mean 15.770; Variance 64.803; scale 0.243

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	49	100.0	56	1	C-terminal Kunitz-type	3.66e+02
2	49	100.0	56	1	C-terminal Kunitz-type	3.66e+02
3	49	100.0	56	1	C-terminal Kunitz-type	3.66e+02
4	49	100.0	58	1	US880256 Seq ID 40.	3.66e+02
5	49	100.0	58	1	C-terminal Kunitz-type	3.66e+02
6	49	100.0	58	1	Human collagen alpha-3	3.66e+02
7	49	100.0	58	1	Human collagen alpha-3	3.66e+02
8	49	100.0	58	1	Human aprotinin-like K	3.66e+02
9	49	100.0	58	1	C-terminal Kunitz-type	3.66e+02
10	49	100.0	58	1	Human Kunitz-type ser1	3.66e+02
11	49	100.0	63	1	Sequence encoded by al	3.66e+02
12	49	100.0	70	1	Sequence of the last 7	3.66e+02
13	49	100.0	100	1	Human TPC2/MBP fusion	3.66e+02
14	49	100.0	111	1	Protein encoded by PKF	3.66e+02
15	49	100.0	113	1	T-cell receptor V-beta	3.66e+02
16	49	100.0	180	1	Streptothricin acetyl	3.66e+02
17	49	100.0	227	1	Human calcium channel	3.66e+02
18	49	100.0	312	1	T-cell antigen recepto	3.66e+02
19	49	100.0	312	1	Portion of a human T-c	3.66e+02
20	49	100.0	370	1	Rat hypothalamic galan	3.66e+02
21	49	100.0	396	1	Protein Male.	3.66e+02
22	49	100.0	396	1	Maltose binding protei	3.66e+02
23	49	100.0	427	1	Human hypothalamic gal	3.66e+02

#### ALIGNMENTS

##### RESULT 1

ID R39676 standard; protein; 56 AA.

AC R39676;

DE 13-JAN-1994 (first entry)

KW Human type VI collagen; preferential inhibition; cathepsin G;

KW neutrophil elastase; protease 3; trypsin; plasmin; kallikrein;

KW pathological proteolysis treatment; polymorphonuclear leucocytes;

KW enzyme release; acute pancreatitis; inflammation; thrombocytopaenia;

KW rheumatoid arthritis; psoriasis; organ preservation;

KW platelet function preservation; domain; example.

OS Homo sapiens.

PN W0931419-A.

PD 22-JUL-1993; DK0002.

PF 07-JAN-1993; DK0002.

PR 07-JAN-1992; WO-DK0005.

PA (NOVO ) NOVO-NORDISK AS.

PI Bjorn SE, Norris F, Norris K, Olsen OH, Petersen LC;

DR WPI: 93-243147/30.

PT Human Kunitz-type protease inhibitor variants - for treating and

PT preventing diseases associated with pathological proteolysis e.g.

PT acute pancreatitis, inflammation, thrombocytopaenia etc.

PS Example; Page 26; 33pp; English.

CC The sequence is that of a variant of the C-terminal Kunitz-type

CC protease inhibitor domain of the alpha3 chain of human type VI

CC collagen. Modification of the Kunitz sequence provides preferential

CC inhibition of e.g. neutrophil elastase, cathepsin G, protease-3,

CC trypsin, plasmin and/or kallikrein. The variant can thus be used to

CC treat pathological proteolysis caused by enzymes released from

CC polymorphonuclear leucocytes, e.g. acute pancreatitis, inflammation,

CC thrombocytopaenia, rheumatoid arthritis, psoriasis, etc., opt.

CC administered with heparin. It can also be used for preservation of

CC organs or platelet function, and possibly for isolating natural cpds.

CC which bind to the specified domain. It has zero net charge and so

CC is less likely to cause kidney damage than aprotinin and, being of

CC human origin is less immunogenic.

SQ Sequence 56 AA;

Query Match 100.0%; Score 49; DB 1; Length 56;

Best Local Similarity 36.4%; Pred. No. 3.66e+02;

Matches 4; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Db 23 PNTKSCAREWY 33

QY 2 PXXXXXXFWY 12

Maltose binding protei 3.66e+02  
Modified p17 antigen o 3.66e+02  
Maltose binding protei 3.66e+02  
ATL protein-maltose bi 3.66e+02  
Invasin-maltose bindin 3.66e+02  
L-proline-4-hydroxylas 3.66e+02  
Tip adhesin protein an 3.66e+02  
Maltose binding protei 3.66e+02  
Maltose binding protei 3.66e+02  
Human calcium channel 3.66e+02  
B. pallidus DNA polyme 3.66e+02  
Chicken leucocytosom 3.66e+02  
Streptokinase/maltose 3.66e+02  
Modified streptokinase 3.66e+02  
Modified streptokinase 3.66e+02  
Streptokinase/maltose 3.66e+02  
Aspergillus flavus mul 3.66e+02  
Human neuronal calcium 3.66e+02  
Human neuronal calcium 3.66e+02  
Sequence of the alpha 3.66e+02  
Human calcium channel 3.66e+02  
Human calcium channel 3.66e+02

24 49 100.0 506 1 W61152  
25 49 100.0 524 1 W43304  
26 49 100.0 535 1 R78524  
27 49 100.0 568 1 R96210  
28 49 100.0 588 1 R96209  
29 49 100.0 659 1 W09294  
30 49 100.0 676 1 R89331  
31 49 100.0 708 1 W06411  
32 49 100.0 711 1 W06412  
33 49 100.0 993 1 R27650  
34 49 100.0 1084 1 W59033  
35 49 100.0 1132 1 R97866  
36 49 100.0 1181 1 W21727  
37 49 100.0 1194 1 W21725  
38 49 100.0 1194 1 W21724  
39 49 100.0 1194 1 W21726  
40 49 100.0 1307 1 R99255  
41 49 100.0 2161 1 R71002  
42 49 100.0 2161 1 R71001  
43 49 100.0 2161 1 R33545  
44 49 100.0 2161 1 W63137  
45 49 100.0 2161 1 W63149



**THIS PAGE BLANK (uspto)**

```
OC Ephydroidea; Drosophilidae; Drosophila.
RN [1]
RA SEQUENCE FROM N.A.
RP ZENG L.-W., COMERON J.M., CHEN B., KREITMAN M.;
RL Genetica 0:0-0(1997).
DR EMBL; AF025808; AAB87893.1; -.
DR HSSP; P06543; IQOC.
DR FLYBASE; FBgn0023244; Dsub\Gad1.
DR PFAM; PF00282; pyridoxal_dec; 2.
FT NON_TER 1
FT NON_TER 370 370
SQ SEQUENCE 370 AA; 41923 MW; D75BDC49 CRC32;

Query Match 100.0%; Score 49; DB 5; Length 370;
Best Local Similarity 40.0%; Pred. No. 1.02e+02;
Matches 4; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 295 PECVNVSWFY 304
QY 2 PXXXXXXFWY 11
|
|
|

RESULT 13
ID O44102 PRELIMINARY; PRT; 370 AA.
AC O44102;
DT 01-JUN-1998 (TREMBlrel. 06, Created)
DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
DE GLUTAMIC ACID DECARBOXYLASE (FRAGMENT).
GN GAD1.
OS Drosophila pseudoobscura (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
RN [1]
RA SEQUENCE FROM N.A.
RP ZENG L.-W., COMERON J.M., CHEN B., KREITMAN M.;
RL Genetica 0:0-0(1997).
DR EMBL; AF025807; AAB87892.1; -.
DR HSSP; P06543; IQOC.
DR FLYBASE; FBgn0023295; Dpse\Gad1.
DR PFAM; PF00282; pyridoxal_dec; 2.
FT NON_TER 1
FT NON_TER 370 370
SQ SEQUENCE 370 AA; 41887 MW; A1EB456F CRC32;

Query Match 100.0%; Score 49; DB 5; Length 370;
Best Local Similarity 40.0%; Pred. No. 1.02e+02;
Matches 4; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 295 PECVNVSWFY 304
QY 2 PXXXXXXFWY 11
|
|
|

RESULT 14
ID Q15227 PRELIMINARY; PRT; 402 AA.
AC Q15227;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
DE PREGNANCY-SPECIFIC BETA-1 GLYCOPROTEIN PRECURSOR.
GN PSG11.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RA SEQUENCE FROM N.A.
RP TISSUE-PLACENTA;
RX MEDLINE; 92017749.
RA CHAN W.Y., ZHENG Q.X., MCMAHON J., TEASE L.A.;
RT "Characterization of new members of the pregnancy-specific beta 1-
glycoprotein family.";

Mol. Cell. Biochem. 106:161-170(1991).
DR EMBL; M94890; AAA60194.1; -.
DR PFAM; PF00047; 1g; 1.
FT SIGNAL 1 35 POTENTIAL.
FT CHAIN 36 402 PREGNANCY-SPECIFIC BETA-1 GLYCOPROTEIN.
SQ SEQUENCE 402 AA; 45336 MW; CF3E4568 CRC32;

Query Match 100.0%; Score 49; DB 4; Length 402;
Best Local Similarity 40.0%; Pred. No. 1.02e+02;
Matches 4; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 59 PQNLPGYFWY 68
QY 2 PXXXXXXFWY 11
|
|
|

RESULT 15
ID Q15237 PRELIMINARY; PRT; 426 AA.
AC Q15237;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
DE PSG11 PRECURSOR.
GN PSG11.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RA SEQUENCE FROM N.A.
RP TISSUE-PLACENTA;
RX MEDLINE; 92256483.
RA BROPHY B.K., MACDONALD R.E., MCLENACHAN P.A., MANSFIELD B.C.;
RT "cDNA sequence of the pregnancy-specific beta 1-glycoprotein-11s (PSG-
11s)."
RL Biochim. Biophys. Acta 1131:119-121(1992).
DR EMBL; M58591; AAA60203.1; -.
DR PFAM; PF00047; 1g; 2.
FT SIGNAL 1 34 POTENTIAL.
FT CHAIN 35 426 POTENTIAL.
SQ SEQUENCE 426 AA; 48332 MW; B9A4E65B CRC32;

Query Match 100.0%; Score 49; DB 4; Length 426;
Best Local Similarity 40.0%; Pred. No. 1.02e+02;
Matches 4; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 59 PQNLPGYFWY 68
QY 2 PXXXXXXFWY 11
|
|
|

Search completed: Sat Apr 15 01:37:50 2000
Job time : 93 secs.
```

DT 01-NOV-1996 (TREMBlrel. 01, Created)  
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
 DE 01-NOV-1998 (TREMBlrel. 08, Last annotation update)  
 DE SIMILAR TO E. COLI GLPT AND UHPT PROTEINS.  
 GN IPGH ORFB.  
 OS Shigella flexneri.  
 OG Plasmid pWR100.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Shigella.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=M9OT-W, SEROTYPE 5;  
 RX MEDLINE; 92167809.  
 RA VENKATESAN M.M., BUYSE J.M., HARTMAN A.B.;  
 RT "sequence variation in two ipaH genes of Shigella flexneri 5 and  
 RL Mol. Microbiol. 5:2435-2445(1991).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=M9OT-W, SEROTYPE 5;  
 RX MEDLINE; 97074644.  
 RA VENKATESAN M.M., ALEXANDER W.A., FERNANDEZ-PRADA C.;  
 RT "A Shigella flexneri invasion plasmid gene, ipgH, with homology to  
 RL IS629 and sequences encoding bacterial sugar phosphate transport  
 RT proteins.";  
 RL Gene 175:23-27(1996).  
 DR EMBL; U28354; AAC44575.1; -.  
 KW Plasmid.  
 SQ SEQUENCE 333 AA; 36475 MW; 0EB41D70 CRC32;  
 Query Match 100.0%; Score 49; DB 2; Length 333;  
 Best Local Similarity 40.0%; Pred. No. 1.02e+02;  
 Matches 4; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
 Db 40 PGAKTLVFWY 49  
 QY 2 PXXXXXXFWY 11  
 RESULT 10  
 ID Q15225 PRELIMINARY; PRT; 351 AA.  
 AC Q15225  
 DT 01-NOV-1996 (TREMBlrel. 01, Created)  
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
 DE 01-NOV-1999 (TREMBlrel. 12, Last annotation update)  
 DE PREGNANCY-SPECIFIC BETA-1-GLYCOPROTEIN (SP1) (FRAGMENT).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 90268037.  
 RA ZHENG Q.X., TEASE L.A., SHUPERT W.L., CHAN W.Y.;  
 RT "Characterization of cDNAs of the human pregnancy-specific beta 1-  
 RT glycoprotein family, a new subfamily of the immunoglobulin gene  
 RT superfamily.";  
 RL Biochemistry 29:2845-2890(1990).  
 DR EMBL; M31126; AAA36509.1; -.  
 DR PFAM; PF00047; 1g; 1.  
 KW Pregnancy.  
 FT NON\_TER  
 SQ SEQUENCE 351 AA; 39874 MW; 51423514 CRC32;  
 Query Match 100.0%; Score 49; DB 4; Length 351;  
 Best Local Similarity 40.0%; Pred. No. 1.02e+02;  
 Matches 4; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
 Db 8 PQNLPGYFWY 17  
 QY 2 PXXXXXXFWY 11  
 RESULT 11

ID 031507 PRELIMINARY; PRT; 365 AA.  
 AC 031507;  
 DT 01-JAN-1998 (TREMBlrel. 05, Created)  
 DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)  
 DE 01-NOV-1998 (TREMBlrel. 08, Last annotation update)  
 DE YEEG PROTEIN.  
 GN YEEG.  
 OS Bacillus subtilis.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
 OC Bacillus/Staphylococcus group; Bacillus.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RX MEDLINE; 98044033.  
 RA KUNST F., OGASAWARA N., MOSZER I., ALBERTINI A.M., ALLONI G.,  
 RA AZEVEDO V., BERTERO M.G., BESSIERES P., BOLOTIN A., BORCHERT S.,  
 RA BORRIS R., BOURSIER L., BRANS A., BRAUN M., BRIGNELL S.C., BRON S.,  
 RA BROUILLET S., BRUSCHI C.V., CALDWELL B., CAPUANO V., CARTER N.M.,  
 RA CHOI S.K., CODANI J.J., CONNERTON I.F., CUMMINGS N.J., DANIEL R.A.,  
 RA DENIZOT F., DEVINE K.M., DUSTERHOFT A., EHRLICH S.D., EMERSON P.T.,  
 RA ENTIAN K.D., ERRINGTON J., FABRET C., FERRARI E., FOULGER D.,  
 RA FRITZ C., FUJITA M., FUJITA Y., FUMA S., GALIZZI A., GALLERON N.,  
 RA GHIM S.Y., GLASER P., GOFFEAU A., GOLIGHTLY E.J., GRANDI G.,  
 RA GUISEPI G., GUY B.J., HAGA K., HAIECH J., HARWOOD C.R., HENAUT A.,  
 RA HILBERT H., HOLSAPPEL S., HOSONO S., HULLO M.F., ITAYA M., JONES L.,  
 RA JORIS B., KARAMATA D., KASAHARA Y., KLAERR-BLANCHARD M., KLEIN C.,  
 RA KOBAYASHI Y., KOETTER P., KONINGSTEIN G., KROGH S., KUMANO M.,  
 RA KURITA K., LAPIDUS A., LARDINOIS S., LAUBER J., LAZAREVIC V.,  
 RA LEE S.M., LEVINE A., LIU H., MASUDA S., MAUEL C., MEDIGUE C.,  
 RA MEDINA N., MELLADO R.P., MIZUNO M., MOESTL D., NAKAI S., NOBACK M.,  
 RA NOONE D., O'REILLY M., OGAWA K., OGAWARA A., OUDEGA B., PARK S.H.,  
 RA PARRO V., POHL T.M., PORTETELLE D., PORWOLLIK S., PRESCOTT A.M.,  
 RA PRESECAN E., PUJIC P., PURNELLE B., RAPOPORT G., REY M., REYNOLDS S.,  
 RA RIEGER M., RIVOLTA C., ROCHA E., ROCHE B., ROSE M., SADAIE Y.,  
 RA SATO T., SCANLAN E., SCHLEICH S., SCHROETER R., SCOFFONE F.,  
 RA SERIGUCHI J., SEKOWSKA A., SERO S.J., SERRO P., SHIN B.S., SOLDI B.,  
 RA SOROKIN A., TACCONI E., TAKAGI T., TAKAHASHI H., TAKEMARU K.,  
 RA TAKEUCHI M., TAMAKOSHI A., TANAKA T., TERPSTRA P., TOGNONI A.,  
 RA TOSATO V., UCHIYAMA S., VANDENBOL M., VANNIER F., VASSAROTTI A.,  
 RA VIARI A., WAMBUTT R., WEDLER E., WEDLER H., WEITZENEGER T.,  
 RA WINTERS P., WIPAT A., YAMAMOTO H., YAMANE K., YASUMOTO K., YATA K.,  
 RA YOSHIDA K., YOSHIKAWA H.F., ZUMSTEIN E., YOSHIKAWA H., DANCHIN A.;  
 RT "The complete genome sequence of the gram-positive bacterium Bacillus  
 RT subtilis.";  
 RL Nature 390:249-256(1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RA KUNST F., OGASAWARA N., YOSHIKAWA H., DANCHIN A.;  
 RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; Z99107; CAB12502.1; -.  
 SQ SEQUENCE 365 AA; 42349 MW; C2CDB5AF CRC32;  
 Query Match 100.0%; Score 49; DB 2; Length 365;  
 Best Local Similarity 40.0%; Pred. No. 1.02e+02;  
 Matches 4; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
 Db 139 PLSDPFAFWY 148  
 QY 2 PXXXXXXFWY 11  
 RESULT 12  
 ID 044103 PRELIMINARY; PRT; 370 AA.  
 AC 044103;  
 DT 01-JUN-1998 (TREMBlrel. 06, Created)  
 DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)  
 DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)  
 DE GLUTAMIC ACID DECARBOXYLASE (FRAGMENT).  
 GN GAD1.  
 OS Drosophila subobscura (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

```
DR PFAM; PF01652; IF4E; 1.  
KW Initiation factor; Protein biosynthesis; RNA-binding;  
KW Mutigene family.  
FT MUTAGEN 63 95 W->A: UNABLE TO BIND CAPPED RNA.  
FT MUTAGEN 95 95 W->A: ABILITY TO BIND CAPPED RNA REDUCED  
TO 40% OF WILD-TYPE.  
FT MUTAGEN 124 126 WED->FAA: UNABLE TO BIND CAPPED RNA.  
FT MUTAGEN 124 124 W->A: ABILITY TO BIND CAPPED RNA REDUCED  
TO LESS THAN 10% OF WILD-TYPE.  
FT MUTAGEN 124 124 W->F: ABILITY TO BIND CAPPED RNA REDUCED  
TO 13% OF WILD-TYPE.  
FT MUTAGEN 125 125 E->A: ABILITY TO BIND CAPPED RNA REDUCED  
TO LESS THAN 10% OF WILD-TYPE.  
FT MUTAGEN 126 126 D->A: SLIGHT REDUCTION IN ABILITY TO BIND  
CAPPED RNA.  
FT MUTAGEN 135 135 W->A: UNABLE TO BIND CAPPED RNA.  
FT MUTAGEN 148 148 W->A: UNABLE TO BIND CAPPED RNA.  
FT MUTAGEN 183 183 W->A: ABILITY TO BIND CAPPED RNA REDUCED  
TO LESS THAN 10% OF WILD-TYPE.  
FT MUTAGEN 183 183 W->F: UNABLE TO BIND CAPPED RNA.  
FT CONFLICT 1 27 MNKFDAKDDSDGDHQNENSTOKD ->  
FT MMTVGTMTIRKKTAHRKI (IN REF. 2).  
SQ SEQUENCE 245 AA; 28362 MW; 226575AC CRC32;  
  
Query Match 100.0%; Score 49; DB 4; Length 245;  
Best Local Similarity 40.0%; Pred. No. 1.02e+02;  
Matches 4; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
  
Db 55 PLQYNYTFWY 64  
|  
|  
|  
Qy 2 PXXXXXXFWY 11  
  
RESULT 6  
ID O88503 PRELIMINARY; PRT; 245 AA.  
AC O88503;  
DT 01-NOV-1998 (TrEMBLrel. 08, Created)  
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)  
DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)  
DE EIF4E-LIKE PROTEIN 4E-LP.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-NIH 3T3 SWISS;  
RA JOSHI B., JAGUS R.;  
RT "Isolation of human and mouse cDNAs encoding novel eIF4E-like  
proteins."  
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AF068116; AAC19373.1; -.  
DR PROSITE: PS00813; IF4E; 1.  
DR PFAM; PF01652; IF4E; 1.  
SQ SEQUENCE 245 AA; 28297 MW; 58683B95 CRC32;  
  
Query Match 100.0%; Score 49; DB 11; Length 245;  
Best Local Similarity 40.0%; Pred. No. 1.02e+02;  
Matches 4; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
  
Db 55 PLQYNYTFWY 64  
|  
|  
|  
Qy 2 PXXXXXXFWY 11  
  
RESULT 7  
ID O76446 PRELIMINARY; PRT; 298 AA.  
AC O76446;  
DT 01-NOV-1998 (TrEMBLrel. 08, Created)  
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)  
DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)  
DE ZK1055.2 PROTEIN.  
GN ZK1055.2.  
OS Caenorhabditis elegans.  
  
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;  
OC Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RX MEDLINE; 94150718.  
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,  
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,  
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,  
RA GARDNER A., GREEN P., HAWKINS J., HILLIER L., JIER M., JOHNSTON L.,  
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,  
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,  
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,  
RA SMAILDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,  
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,  
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;  
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
elegans.";  
RL Nature 368:32-38(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RA GEISEL C., BRADSHAW H.;  
RT "The sequence of C. elegans cosmid ZK1055.";  
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RA WATERSTON R.;  
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AF068721; AAC19260.1; -.  
SQ SEQUENCE 298 AA; 34973 MW; 17907754 CRC32;  
  
Query Match 100.0%; Score 49; DB 5; Length 298;  
Best Local Similarity 40.0%; Pred. No. 1.02e+02;  
Matches 4; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
  
Db 229 PSQMGRDFWY 238  
|  
|  
|  
Qy 2 PXXXXXXFWY 11  
  
RESULT 8  
ID O54089 PRELIMINARY; PRT; 310 AA.  
AC O54089;  
DT 01-JUN-1998 (TrEMBLrel. 06, Created)  
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)  
DT 01-JUN-1998 (TrEMBLrel. 06, Last annotation update)  
DE CYTOCHROME B558/566, SUBUNIT B.  
GN CBSB.  
OS Sulfolobus acidocaldarius.  
OC Archaea; Crenarchaeota; Sulfolobales; Sulfolobus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-DSM 639;  
RA HETTMANN T., SCHMIDT C.L., ANEMUELLER S., ZAEHRINGER U., MOLL H.,  
RA PETERSEN A., SCHAEFER G.;  
RL J. Biol. Chem. 0:0-0(0).  
DR EMBL: Y10108; CAA71196.1; -.  
SQ SEQUENCE 310 AA; 35150 MW; 83DF475A CRC32;  
  
Query Match 100.0%; Score 49; DB 1; Length 310;  
Best Local Similarity 40.0%; Pred. No. 1.02e+02;  
Matches 4; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
  
Db 94 PNYNNSPFWY 103  
|  
|  
|  
Qy 2 PXXXXXXFWY 11  
  
RESULT 9  
ID Q54147 PRELIMINARY; PRT; 333 AA.  
AC Q54147;
```

DE 01-NOV-1999 (Tremblrel. 12, Last annotation update)  
DE CARCINOEMBRYONIC ANTIGEN 3 (PREGNANCY-SPECIFIC GLYCOPROTEIN)  
DE (FRAGMENT).  
GN CEA3.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 92345715.  
RA RUDERT F., SAUNDERS A.M., REBSTOCK S., THOMPSON J.A., ZIMMERMANN W.;  
RT "Characterization of murine carcinoembryonic antigen gene family  
RT members".  
RL Mamm. Genome 3:262-273(1992).  
DR EMBL; M83346; AAA39917.1; -.  
DR MGD; MGI:88368; Cea3.  
KW Pregnancy.  
FT NON\_TER 209 209  
SQ SEQUENCE 209 AA; 23824 MW; D8D86073 CRC32;  
Query Match 100.0%; Score 49; DB 11; Length 209;  
Best Local Similarity 40.0%; Pred. No. 1.02e+02;  
Matches 4; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
Db 180 PKYQSLFWY 189  
QY | | |  
2 PXXXXXXFWY 11  
RESULT 3  
ID Q15461 PRELIMINARY; PRT; 236 AA.  
AC Q15461;  
DT 01-NOV-1996 (Tremblrel. 01, Created)  
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)  
DT 01-MAY-1999 (Tremblrel. 10, Last annotation update)  
DE PREGNANCY-SPECIFIC BETA-1 GLYCOPROTEIN-11 (FRAGMENT).  
GN PSG11.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-BLOOD;  
RX MEDLINE; 95104846.  
RA MCLENACHAN P.A., RUTHERFORD K.J., BEGGS K.T., SIMS S.E.,  
RA MANSFIELD B.C.;  
RT "Characterization of the PSG11 gene".  
RL Genomics 22:356-363(1994).  
DR EMBL; U08196; AAA78805.1; -.  
DR EMBL; U08194; AAA78805.1; JOINED.  
DR EMBL; U08195; AAA78805.1; JOINED.  
KW Pregnancy.  
FT NON\_TER 236 236  
SQ SEQUENCE 236 AA; 26816 MW; 9E47AD86 CRC32;  
Query Match 100.0%; Score 49; DB 4; Length 236;  
Best Local Similarity 40.0%; Pred. No. 1.02e+02;  
Matches 4; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
Db 59 PQNLPGYFWY 68  
QY | | |  
2 PXXXXXXFWY 11  
RESULT 4  
ID Q13178 PRELIMINARY; PRT; 240 AA.  
AC Q13178;  
DT 01-NOV-1996 (Tremblrel. 01, Created)  
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)  
DT 01-NOV-1998 (Tremblrel. 08, Last annotation update)  
DE PREGNANCY-SPECIFIC GLYCOPROTEIN 11.  
GN PSG11s.  
OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-LIVER;  
RX MEDLINE; 95314639.  
RA TEGLUND S., ZHOU G.Q., HAMMARSTROM S.;  
RT "Characterization of cDNA encoding novel pregnancy-specific  
RT glycoprotein variants".  
RL Biochem. Biophys. Res. Commun. 211:656-664(1995).  
DR EMBL; U25987; AAA75298.1; -.  
DR PFAM; PF00047; Ig; 1.  
DR Pregnancy.  
SQ SEQUENCE 240 AA; 27004 MW; 40A35D60 CRC32;  
Query Match 100.0%; Score 49; DB 4; Length 240;  
Best Local Similarity 40.0%; Pred. No. 1.02e+02;  
Matches 4; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
Db 59 PQNLPGYFWY 68  
QY | | |  
2 PXXXXXXFWY 11  
RESULT 5  
ID O60573 PRELIMINARY; PRT; 245 AA.  
AC O60573; 075349;  
DT 01-AUG-1998 (Tremblrel. 07, Created)  
DT 01-AUG-1998 (Tremblrel. 07, Last sequence update)  
DT 01-NOV-1999 (Tremblrel. 12, Last annotation update)  
DE EUKARYOTIC TRANSLATION INITIATION FACTOR 4E HOMOLOGOUS PROTEIN (MRNA  
DE CAP-BINDING PROTEIN 4EHP) (4E HOMOLOGOUS PROTEIN) (IF4E-LIKE PROTEIN  
DE 4E-LP).  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.  
RN [1]  
RP SEQUENCE FROM N.A., 3D-STRUCTURE MODELING, AND MUTAGENESIS.  
RC TISSUE-FOLLICLE;  
RX MEDLINE; 98250763.  
RA ROM E., KIM H.C., GINGRAS A.-C., MARCOTRIGIANO J., FAVRE D., OLSEN H.,  
RA BURLEY S.K., SONENBERG N.;  
RT "Cloning and characterization of 4EHP, a novel mammalian eIF4E-related  
RT cap-binding protein".  
RL J. Biol. Chem. 273:13104-13109(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE-BLOOD;  
RX MEDLINE; 98318631.  
RA MAO M., FU G., WU J.-S., ZHANG Q.-H., ZHOU J., KAN L.-X., HUANG Q.-H.,  
RA HE K.-L., GU B.-W., HAN Z.-G., SHEN Y., GU J., YU Y.-P., XU S.-H.,  
RA WANG Y.-X., CHEN S.-J., CHEN Z.;  
RT "Identification of genes expressed in human CD34(+) hematopoietic  
RT stem/progenitor cells by expressed sequence tags and efficient full-  
RT length cDNA cloning".  
RL Proc. Natl. Acad. Sci. U.S.A. 95:8175-8180(1998).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC TISSUE-BREAST;  
RA JOSHI B., NORRIS K.K., JAGUS R.;  
RT "Isolation of human and mouse cDNAs encoding novel eIF4E-like  
RT proteins".  
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.  
CC -|- FUNCTION: RECOGNIZES AND BINDS THE 7-METHYLGUANOSINE-CONTAINING  
CC MRNA "CAP" DURING AN EARLY STEP IN THE INITIATION OF PROTEIN  
CC SYNTHESIS AND FACILITATES RIBOSOME BINDING BY INDUCING THE  
CC UNWINDING OF THE MRNA SECONDARY STRUCTURES (BY SIMILARITY).  
CC -|- SUBCELLULAR LOCATION: CYTOPLASMIC.  
CC -|- SIMILARITY: TO EIF4E.  
DR EMBL; AF047695; AAC18565.1; -.  
DR EMBL; AF038957; AAC39871.1; -.  
DR EMBL; AF068117; AAC19374.1; -.  
DR PROSITE; PS00813; IF4E, 1.



Site  
PAGE BLANK (uspto)

QY 1  
2 PXXXXXXFWY 11

RESULT 14  
ID VGLY\_TACVT STANDARD; PRT; 483 AA.  
AC P31841;  
DT 01-JUL-1993 (Rel. 26, Created)  
DT 01-JUL-1993 (Rel. 26, Last sequence update)  
DT 01-JUL-1993 (Rel. 26, Last annotation update)  
DE GLYCOPROTEIN POLYPROTEIN PRECURSOR [CONTAINS: GLYCOPROTEINS G1 AND G2].  
DE G2].  
GN GPC.  
OS Tacaribe virus (strain V5).  
OC Viruses; ssRNA negative-strand viruses; Arenaviridae; Arenavirus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 91341495.  
RA ALLISON L.M.C., SALTER M.W.A.P., KIGUWA S., HOWARD C.R.;  
RT "Analysis of the glycoprotein gene of Tacaribe virus and neutralization-resistant variants";  
RL J. Gen. Virol. 72:2025-2029(1991).  
CC -|- SIMILARITY: BELONGS TO THE ARENAVIRUSES GPC PROTEIN FAMILY.  
DR PIR; JQ1454; VGXPT5.  
DR PFAM; PF00798; Arena-glycoprot; 1.  
KW Polyprotein; Glycoprotein; Envelope protein.  
FT CHAIN 1 246 SURFACE GLYCOPROTEIN G1.  
FT CHAIN 247 483 SURFACE GLYCOPROTEIN G2.  
FT CARBOHYD 83 83 POTENTIAL.  
FT CARBOHYD 95 95 POTENTIAL.  
FT CARBOHYD 164 164 POTENTIAL.  
FT CARBOHYD 176 176 POTENTIAL.  
FT CARBOHYD 355 355 POTENTIAL.  
FT CARBOHYD 363 363 POTENTIAL.  
FT CARBOHYD 380 380 POTENTIAL.  
FT CARBOHYD 385 385 POTENTIAL.  
SQ SEQUENCE 483 AA; 55598 MW; DA32E4FD CRC32;

Query Match 100.0%; Score 49; DB 1; Length 483;  
Best Local Similarity 40.0%; Pred.No. 4.51e+01;  
Matches 4; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 352 PYCNYTRFWY 361  
1  
2 PXXXXXXFWY 11

RESULT 15  
ID VGLY\_TACVT STANDARD; PRT; 483 AA.  
AC P31840;  
DT 01-JUL-1993 (Rel. 26, Created)  
DT 01-JUL-1993 (Rel. 26, Last sequence update)  
DT 01-OCT-1993 (Rel. 27, Last annotation update)  
DE GLYCOPROTEIN POLYPROTEIN PRECURSOR [CONTAINS: GLYCOPROTEINS G1 AND G2].  
DE G2].  
GN GPC.  
OS Tacaribe virus (strain TRVL 11598).  
OC Viruses; ssRNA negative-strand viruses; Arenaviridae; Arenavirus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 91341495.  
RA ALLISON L.M.C., SALTER M.W.A.P., KIGUWA S., HOWARD C.R.;  
RT "Analysis of the glycoprotein gene of Tacaribe virus and neutralization-resistant variants";  
RL J. Gen. Virol. 72:2025-2029(1991).  
CC -|- SIMILARITY: BELONGS TO THE ARENAVIRUSES GPC PROTEIN FAMILY.  
DR PIR; JQ1453; VGXPTV.  
DR PFAM; PF00798; Arena.glycoprot; 1.  
KW Polyprotein; Glycoprotein; Envelope protein.  
FT CHAIN 1 246 SURFACE GLYCOPROTEIN G1.  
FT CHAIN 247 483 SURFACE GLYCOPROTEIN G2.  
FT CARBOHYD 83 83 POTENTIAL.  
FT CARBOHYD 95 95 POTENTIAL.

FT CARBOHYD 164 164 POTENTIAL.  
FT CARBOHYD 176 176 POTENTIAL.  
FT CARBOHYD 355 355 POTENTIAL.  
FT CARBOHYD 363 363 POTENTIAL.  
FT CARBOHYD 380 380 POTENTIAL.  
FT CARBOHYD 385 385 POTENTIAL.  
SQ SEQUENCE 483 AA; 55602 MW; 03B3671E CRC32;

Query Match 100.0%; Score 49; DB 1; Length 483;  
Best Local Similarity 40.0%; Pred.No. 4.51e+01;  
Matches 4; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 352 PYCNYTRFWY 361  
1  
2 PXXXXXXFWY 11

Search completed: Sat Apr 15 01:35:59 2000  
Job time : 41 secs.



[4]  
RN X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).  
RP MEDLINE; 98188227.  
RA WILCE M.C., BOND C.S., DIXON N.E., FREEMAN H.C., GUSS J.M.,  
RA LILLEY P.E., WILCE J.A.;  
RT Structure and mechanism of a proline-specific aminopeptidase from  
RT Escherichia coli.;  
RL Proc. Natl. Acad. Sci. U.S.A. 95:3472-3477(1998).  
CC -1- CATALYTIC ACTIVITY: RELEASES ANY N-TERMINAL AMINO ACID, INCLUDING  
CC PROLINE, THAT IS LINKED WITH PROLINE, EVEN FROM A DIPEPTIDE OR  
CC TRIPEPTIDE.  
CC -1- COFACTOR: REQUIRES ZINC. THE INACTIVE APOENZYME CAN BE ACTIVATED  
CC BY THE ADDITION OF MANGANESE AS WELL AS ZINC.  
CC -1- SUBUNIT: HOMOTETRAMER.  
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M24B.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; D00398; BAA00299.1; -  
DR EMBL; D90281; BAA14325.1; -  
DR EMBL; U28377; AAG69076.1; -  
DR EMBL; AE000374; AAC75946.1; -  
DR PIR; JX0067; DPECP.  
DR PIR; JQ0843; JQ0843.  
DR PIR; B47020; B47020.  
DR PDB; 1A29; 06-APR-99.  
DR PDB; 1JAW; 06-APR-99.  
DR PDB; 1A16; 06-APR-99.  
DR ECOGENE; EGI0697; PEP.  
DR PROSITE; PS00491; PROLINE\_PEPTIDASE; 1.  
DR PFAM; PF00357; Peptidase\_M24; 1.  
KW Hydrolase; Aminopeptidase; Manganese; Zinc; 3D-structure.  
FT INIT MET 0  
SQ SEQUENCE 440 AA; 49684 MW; E72D265A CRC32;

Query Match 100.0%; Score 49; DB 1; Length 440;  
Best Local Similarity 40.0%; Pred. No. 4.51e+01;  
Matches 4; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 42 PYRQNSDEWY 51  
QY 2 PXXXXXXFWY 11

RESULT 12  
ID VGLX JUNIN STANDARD; PRT; 481 AA.  
AC P26313;  
DT 01-MAY-1992 (Rel. 22, Created)  
DT 01-MAY-1992 (Rel. 22, Last sequence update)  
DT 01-OCT-1993 (Rel. 27, Last annotation update)  
DE GLYCOPROTEIN POLYPROTEIN PRECURSOR [CONTAINS: GLYCOPROTEINS G1 AND  
DE G2].  
GN GPC.  
OS Junin arenavirus.  
OC Viruses; ssRNA negative-strand viruses; Arenaviridae; Arenavirus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-MC2;  
RX MEDLINE; 91374010.  
RA GHIRINGHELLI P.D., RIVERA-POMAR R.V., LOZANO M.E., GRAU O.,  
RA ROMANOWSKI V.;  
RT "Molecular organization of Junin virus S RNA: complete nucleotide  
RT sequence, relationship with other members of the Arenaviridae and  
RT unusual secondary structures";  
RL J. Gen. Virol. 72:2129-2141(1991).  
CC -1- SIMILARITY: BELONGS TO THE ARENAVIRUSES GPC PROTEIN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; D10072; BAA00964.1; -  
DR PIR; J70978; VGXPTV.  
DR PFAM; PF00798; Arena\_glycoprot; 1.  
KW Polyprotein; Glycoprotein; Envelope protein.  
FT CHAIN 1 244  
FT CHAIN 245 481  
FT CARBOHYD 91 91  
FT CARBOHYD 101 101  
FT CARBOHYD 162 162  
FT CARBOHYD 174 174  
FT CARBOHYD 353 353  
FT CARBOHYD 361 361  
FT CARBOHYD 378 378  
FT CARBOHYD 383 383  
SQ SEQUENCE 481 AA; 55118 MW; 920D9226 CRC32;

Query Match 100.0%; Score 49; DB 1; Length 481;  
Best Local Similarity 40.0%; Pred. No. 4.51e+01;  
Matches 4; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 350 PYCNTKFWY 359  
QY 2 PXXXXXXFWY 11

RESULT 13  
ID VGLX TACV7 STANDARD; PRT; 482 AA.  
AC P31842;  
DT 01-JUL-1993 (Rel. 26, Created)  
DT 01-JUL-1993 (Rel. 26, Last sequence update)  
DT 01-OCT-1993 (Rel. 27, Last annotation update)  
DE GLYCOPROTEIN POLYPROTEIN PRECURSOR [CONTAINS: GLYCOPROTEINS G1 AND  
DE G2].  
GN GPC.  
OS Tacaribe virus (strain V7).  
OC Viruses; ssRNA negative-strand viruses; Arenaviridae; Arenavirus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 91341495.  
RA ALLISON L.M.C., SALTER M.W.A.P., KIGUWA S., HOWARD C.R.;  
RT "Analysis of the glycoprotein gene of Tacaribe virus and  
RT neutralization-resistant variants";  
RL J. Gen. Virol. 72:2025-2029(1991).  
CC -1- SIMILARITY: BELONGS TO THE ARENAVIRUSES GPC PROTEIN FAMILY.

DR PIR; JQ1455; VGXPTV.  
DR PFAM; PF00798; Arena\_glycoprot; 1.  
KW Polyprotein; Glycoprotein; Envelope protein.  
FT CHAIN 1 246  
FT CHAIN 247 482  
FT CARBOHYD 83 83  
FT CARBOHYD 95 95  
FT CARBOHYD 164 164  
FT CARBOHYD 176 176  
FT CARBOHYD 354 354  
FT CARBOHYD 362 362  
FT CARBOHYD 379 379  
FT CARBOHYD 384 384  
SQ SEQUENCE 482 AA; 55607 MW; 1614401A CRC32;

Query Match 100.0%; Score 49; DB 1; Length 482;  
Best Local Similarity 40.0%; Pred. No. 4.51e+01;  
Matches 4; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 351 PYCNTKFWY 360

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL: X17610; CAA35612.1; -;  
EMBL: M34421; AAA52605.1; -;  
EMBL: M34481; AAA74512.1; -;  
EMBL: AC005392; AAC28916.1; -;  
EMBL: M38243; AAA63252.1; -;  
EMBL: U04325; AAA78266.1; -;  
EMBL: M176398; -;  
PFAM: PF00047; 1g; 2;  
KW Immunoglobulin domain; Glycoprotein; Signal; Multigene family.  
FT SIGNAL 1 34  
FT CHAIN 35 426  
FT FT  
FT DOMAIN 35 144  
FT DOMAIN 162 224  
FT DOMAIN 255 317  
FT DOMAIN 347 401  
FT SITE 127 129  
FT SITE 169 217  
FT DISULFID 262 310  
FT DISULFID 262 310  
FT DISULFID 354 394  
FT CARBOHYD 104 104  
FT CARBOHYD 111 111  
FT CARBOHYD 199 199  
FT CARBOHYD 268 268  
FT CARBOHYD 303 303  
FT CARBOHYD 387 387  
FT CARBOHYD 203 203  
FT CONFLICT 203 203  
FT CONFLICT Y -> C (IN REF. 1).  
SQ SEQUENCE 426 AA; 48272 MW; 6E525205 CRC32;  
  
Query Match 100.0%; Score 49; DB 1; Length 426;  
Best Local Similarity 40.0%; Pred. No. 4.51e+01;  
Matches 4; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
  
Db 59 P0NLPYGFYFW 68  
QY 2 PXXXXXXFWY 11  
-----  
RESULT 10  
ID AMPP\_HAEIN STANDARD; PRT; 430 AA.  
AC P44881;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 01-NOV-1995 (Rel. 32, Last annotation update)  
DE XAA-PRO AMINOPEPTIDASE (EC 3.4.11.9) (X-PRO AMINOPEPTIDASE)  
DE (AMINOPEPTIDASE P II) (APP-II) (AMINOACYLPROLINE AMINOPEPTIDASE).  
GN PEPP OR HI0816.  
OS Haemophilus influenzae.  
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
OC Haemophilus.  
CC [1]  
RN SEQUENCE FROM N.A.  
RC STRAIN-RD / KW20;  
RX MEDLINE: 95350630.  
RA KERLAVAGE A.R., BULT C.J., TOMB J.-F., DOUGHERTY B.A., MERRICK J.N.,  
RA SCOTT J.D., SHIRLEY R., LIU L.-I., GLODEK A., KELLEY J.M.,  
RA WEIDMAN J.F., PHILLIPS C.A., SPRIGGS T., HEDBLUM E., COTTON M.D.,  
RA UTTERBACK T.R., HANNA M.C., NGUYEN D.T., SAUDEK D.M., BRANDON R.C.,  
RA FINE L.D., FRITCHMAN J.L., FUHRMANN J.L., GEOGHAGEN N.S.M.,  
RA GNEHM C.L., McDONALD L.A., SMALL K.V., FRASER C.M., SMITH H.O.,  
RA VENTER J.C.;  
RT "Whole-genome random sequencing and assembly of Haemophilus

RT Influenzae Rd.";  
RL Science 269:496-512(1995).  
CC -1- CATALYTIC ACTIVITY: RELEASES ANY N-TERMINAL AMINO ACID, INCLUDING  
CC PROLINE THAT IS LINKED WITH PROLINE, EVEN FROM A DIPEPTIDE OR  
CC TRIPEPTIDE.  
CC -1- COFACTOR: REQUIRES ZINC. THE INACTIVE APOENZYME CAN BE ACTIVATED  
CC BY THE ADDITION OF MANGANESE AS WELL AS ZINC (BY SIMILARITY).  
CC -1- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M24B.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: U32764; AAC22475.1; -;  
CC TIGR; HI0816; -;  
CC PROSITE: PS00491; PROLINE PEPTIDASE; 1.  
CC PFAM: PF00557; Peptidase\_M24; 1.  
CC KW Hydrolase; Amino-peptidase; Manganese; Zinc.  
CC SQ SEQUENCE 430 AA; 49261 MW; AF7D468C CRC32;  
  
Query Match 100.0%; Score 49; DB 1; Length 430;  
Best Local Similarity 40.0%; Pred. No. 4.51e+01;  
Matches 4; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
  
Db 48 PF00557FWY 57  
QY 2 PXXXXXXFWY 11  
-----  
RESULT 11  
ID AMPP\_ECOLI STANDARD; PRT; 440 AA.  
AC P15034;  
DT 01-APR-1990 (Rel. 14, Created)  
DT 01-APR-1990 (Rel. 14, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE XAA-PRO AMINOPEPTIDASE (EC 3.4.11.9) (X-PRO AMINOPEPTIDASE)  
DE (AMINOPEPTIDASE P II) (APP-II) (AMINOACYLPROLINE AMINOPEPTIDASE).  
GN PEPP.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Escherichia.  
CC [1]  
RN SEQUENCE FROM N.A., SEQUENCE OF 1-20, AND CHARACTERIZATION.  
RC STRAIN-HB101;  
RX MEDLINE: 89278065.  
RA YOSHIMOTO T., TONE H., HONDA T., OSATOMI K., KOBAYASHI R., TSURU D.;  
RT "Sequencing and high expression of aminopeptidase P gene from  
RT Escherichia coli HB101."  
RL J. Biochem. 105:412-416(1989).  
RN [2]  
RN SEQUENCE FROM N.A.  
RC STRAIN-HB101;  
RX MEDLINE: 93054351.  
RA NAKAHIGASHI K., MIYAMOTO K., NISHIMURA K., INOKUCHI H.;  
RT "Isolation and characterization of a light-sensitive mutant of  
RT Escherichia coli K-12 with a mutation in a gene that is required for  
RT the biosynthesis of ubiquinone."  
RL J. Bacteriol. 174:7352-7359(1992).  
RN [3]  
RN SEQUENCE FROM N.A.  
RC STRAIN-K12 / MGL1655;  
RX MEDLINE: 97426617.  
RA BLATTNER F.R., PLUNKETT G. III, BLOCH C.A., PERNA N.T., BURLAND V.,  
RA RILEY M., COLLADO-VIDES J., GLASNER F.D., RODE C.K., MAYHEW G.F.,  
RA GREGOR J., DAVIS N.W., KIRKPATRICK H.A., GOEDEN M.A., ROSE D.J.,  
RA MAU B., SHAO Y.;  
RT "The complete genome sequence of Escherichia coli K-12."  
RL Science 277:1453-1474(1997).

FT BINDING 134 134 ATP (BY SIMILARITY).  
FT ACT\_SITE 235 235 BY SIMILARITY.  
FT CONFLICT 125 146 LSATGMAVLKIKTFHGGTQAPD ->  
CPPEWPCSRPETAAPRPPT (IN REF. 2).  
FT CONFLICT 287 287 P -> S (IN REF. 2).  
FT CONFLICT 293 328 GRGNAGVPGVIGIOLINIMVERMERFVNHTW ->  
DRVDRPDSAOHHGGTGGAGIREPHL (IN REF. 2).  
FT CONFLICT 357 358 LA -> WP (IN REF. 2).  
SQ SEQUENCE 391 AA; 44003 MW; 68E60EA0 CRC32;  
Query Match 100.0%; Score 49; DB 1; Length 391;  
Best Local Similarity 40.0%; Pred. No. 4.51e+01;  
Matches 4; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
Db 261 PGETMTSFY 270  
QY 2 PXXXXXXFWY 11  
RESULT 8  
ID AAPQ\_RHLV STANDARD; PRT; 400 AA.  
AC Q52813;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 15-JUL-1998 (Rel. 36, Last annotation update)  
DE GENERAL L-AMINO ACID TRANSPORT PERMEASE PROTEIN AAPQ.  
GN AAPQ.  
OS Rhizobium leguminosarum (biovar viciae).  
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
OC Rhizobiaceae; Rhizobium.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-3841;  
RX MEDLINE; 97054013.  
RA WALSHAW D.L., POOLE P.S.;  
RT "The general L-amino acid permease of Rhizobium leguminosarum is an ABC uptake system that also influences efflux of solutes.";  
RT Mol. Microbiol. 21:1239-1252(1996).  
CC -1- FUNCTION: PART OF A BINDING-PROTEIN-DEPENDENT TRANSPORT SYSTEM FOR L-AMINO ACIDS. AFFECTS THE UPTAKE AS WELL AS EFFLUX OF THESE AMINO ACIDS. PROBABLY RESPONSIBLE FOR THE TRANSLLOCATION OF THE SUBSTRATE ACROSS THE MEMBRANE.  
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE (PROBABLE).  
CC -1- SIMILARITY: WITH INTEGRAL MEMBRANE COMPONENTS OF OTHER BINDING-PROTEIN-DEPENDENT TRANSPORT SYSTEMS. BELONGS TO THE HISMQ SUBFAMILY. SEEMS TO BE THE ORTHOLOG OF E.COLI YHDX.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation. The European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; X82596; CAA57934.1;  
DR PROSITE; PS00402; BPD\_TRANSP\_INN\_MEMBER; 1.  
DR PFAM; PF00528; BPD\_transp; 1.  
KW Transport; Amino-acid transport; Transmembrane; Inner membrane.  
FT TRANSMEM 29 49 POTENTIAL.  
FT TRANSMEM 100 120 POTENTIAL.  
FT TRANSMEM 142 162 POTENTIAL.  
FT TRANSMEM 188 208 POTENTIAL.  
FT TRANSMEM 225 245 POTENTIAL.  
FT TRANSMEM 264 284 POTENTIAL.  
FT TRANSMEM 341 361 POTENTIAL.  
FT TRANSMEM 367 387 POTENTIAL.  
SQ SEQUENCE 400 AA; 43296 MW; 5EE9C75E CRC32;

Query Match 100.0%; Score 49; DB 1; Length 400;  
Best Local Similarity 40.0%; Pred. No. 4.51e+01;  
Matches 4; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 143 PPLLVIFFWY 152  
QY 2 PXXXXXXFWY 11  
RESULT 9  
ID PSGB\_HUMAN STANDARD; PRT; 426 AA.  
AC Q0087;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 15-JUL-1999 (Rel. 38, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE PREGNANCY-SPECIFIC BETA-1-GLYCOPROTEIN 11 PRECURSOR (PSBG-11)  
DE (PREGNANCY-SPECIFIC BETA-1-GLYCOPROTEIN 11) (PREGNANCY-SPECIFIC BETA-1-GLYCOPROTEIN B) (PS34) (PSG7).  
GN PSG11.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-PLACENTA;  
RX MEDLINE; 90212666.  
RA ARAKAWA F., KUROKI M., MISUMI Y., MATSUO Y., MATSUOKA Y.;  
RT "The nucleotide and deduced amino acid sequences of a cDNA encoding a new species of pregnancy-specific beta 1-glycoprotein (PS beta G).";  
RL Biochim. Biophys. Acta 1048:303-305(1990).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE-PLACENTA;  
RX MEDLINE; 90256167.  
RA STREYDIO C., SWISSENS S., GEORGES M., SZPIRER C., VASSART G.;  
RT "Structure, evolution and chromosomal localization of the human pregnancy-specific beta 1 glycoprotein gene family.";  
RL Genomics 6:579-592(1990).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC TISSUE-LIVER;  
RX MEDLINE; 90226362.  
RA KHAN W.N., HAMMARSTRÖM S.;  
RT "Identification of a new carcinoembryonic antigen (CEA) family member in human fetal liver -- cloning and sequence determination of pregnancy-specific glycoprotein 7.";  
RL Biochem. Biophys. Res. Commun. 168:214-225(1990).  
RN [4]  
RP SEQUENCE FROM N.A.  
RC TISSUE-SPERM;  
RA LAWERDIN J.E., MCCREADY P.M., SKOWRONSKI E., ADAMSON A.W., BURKHART-SCHULTZ K., GORDON L., KYLE A., RAMIREZ M., STILLWAGEN S., PHAN H., VELASCO N., DO L., REGALA W., TERRY A., GARNES J., DANGANAN L., POUNDSTONE P., CHRISTENSEN M., GEORGESCU A., AVILA J., LIU S., ATTIX C., ANDREISE T., TRANKHEIM M., AMICO-KELLER G., COFFIELD J., DUARTE S., LUCAS S., BRUCE R., THOMAS P., QUAN G., KONNILLER B., ARELLANO A., MONTGOMERY M., OW D., NOLAN M., TRONG S., KOBAYASHI A., OLSEN A.S., CARRANO A.V.;  
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.  
RN [5]  
RP SEQUENCE OF 23-143 FROM N.A.  
RC TISSUE-BLOOD;  
RA BEGGS K.T., MCLENACHAN T., MANSFIELD B.;  
RL Submitted (AUG-1991) to the EMBL/GenBank/DBJ databases.  
RN [6]  
RP SEQUENCE OF 416-426 FROM N.A.  
RX MEDLINE; 95104846.  
RA MCLENACHAN P.A., RUTHERFORD K.J., BEGGS K.T., SIMS S.E., MANSFIELD B.C.;  
RT "Characterization of the PSG11 gene.";  
RL Genomics 22:356-363(1994).  
CC -1- DEVELOPMENTAL STAGE: PSBG ARE PRODUCED IN HIGH QUANTITY DURING PREGNANCY.  
CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. CONTAINS 3 C2-LIKE AND ONE V-LIKE DOMAINS. BELONGS TO THE CARCINOEMBRYONIC ANTIGEN SUBFAMILY.

```

CC entitles requires a license agreement (See http://www.isb-sib.ch/announcement)
CC or send an email to licensee@isb-sib.ch).
-----
CC EMBL; U18997; AAA58073.1; ALT_INIT.
DR EMBL; AE000405; AAC76301.1; ALT_INIT.
DR ECOCENE; EG12835; YHDX
DR PROSITE; PS00402; BPD_TRANSF_INN_MEMBER; 1.
DR PFAM; PF00528; BPD_transp; 1.
KW Hypothetical protein; Transport; Amino-acid transport; Transmembrane;
KW Inner membrane.
FT TRANSMEM 21 41 POTENTIAL.
FT TRANSMEM 92 112 POTENTIAL.
FT TRANSMEM 128 148 POTENTIAL.
FT TRANSMEM 180 200 POTENTIAL.
FT TRANSMEM 219 239 POTENTIAL.
FT TRANSMEM 256 276 POTENTIAL.
FT TRANSMEM 333 353 POTENTIAL.
SQ SEQUENCE 362 AA; 39749 MW; 15809238 CRC32;

Query Match 100.0%; Score 49; DB 1; Length 362;
Best Local Similarity 40.0%; Pred. No. 4.51e-01;
Matches 4; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 135 PPLLQIFFWY I44
|
|||
QQ 2 PXXXXXXFWY I1

RESULT 7
ID KR14_HSVI1 STANDARD; PRT; 391 AA.
AC PI5443;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE GENE 14 PROTEIN KINASE [EC 2.7.1.-].
GN 14.
OS Ictalurid herpesvirus 1 (Channel catfish virus) (CCV).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC unclassified Herpesviridae.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-AUBURN 1;
RX MEDLINE; 92087490.
RA DAVISON A.J.;
RT "Channel catfish virus: a new type of herpesvirus.";
RL Virology 186:9-14(1992).
RN [2]
RP SEQUENCE OF 114-391 FROM N.A.
RX MEDLINE; 90272416.
RA LACASA M.;
RT "A protein kinase-related gene within the channel catfish herpesvirus genome.";
RL Nucleic Acids Res. 18:3050-3050(1990).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to licensee@isb-sib.ch).
CC -----
CC EMBL; W751136; AAA88117.1; -
DR EMBL; W751136; AAA88119.1; -
DR EMBL; X15978; CAA34100.1; ALT_INIT.
DR PIR; F36787; TVBE11.
DR PIR; S14686; S14686.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; FALSE_NEG.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR PFAM; PF00069; pkinase; 2.
DR Transfrase; Serine/threonine-protein kinase; ATP-binding.
KW DOMAIN 109 391
FT PROTEIN KINASE.

```

DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE HYPOTHETICAL 20.1 KD PROTEIN YCF52 (ORF174).  
 GN YCF52.  
 OS Porphyra purpurea.  
 OG Chloroplast.  
 OC Eukaryota; Rhodophyta; Bangiophyceae; Bangiales; Bangiaceae; Porphyra.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-AVONPORT;  
 RA REITH M.E., MUNHOLLAND J.;  
 RT "complete nucleotide sequence of the Porphyra purpurea chloroplast genome."  
 RL Plant Mol. Biol. Rep. 13:333-335(1995).  
 CC -|- SIMILARITY: BELONGS TO THE YCF52 FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; U38804; AAC08078.1; -;  
 DR PFM; PF00583; Acetyltransf; 1.  
 KW Hypothetical protein; Chloroplast.  
 SQ SEQUENCE 174 AA; 20099 MW; E39AB564 CRC32;  
 Query Match 100.0%; Score 49; DB 1; Length 174;  
 Best Local Similarity 40.0%; Pred. No. 4.51e+01;  
 Matches 4; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
 Db 163 PDGVKGMFWY 172  
 QY 2 PXXXXXXFWY 11  
 RESULT 3  
 ID Y52L.PROMA STANDARD; PRT; 180 AA.  
 AC Q51893;  
 DT 15-JUL-1999 (Rel. 38, Created)  
 DT 15-JUL-1999 (Rel. 38, Last sequence update)  
 DE YCF52-LIKE PROTEIN.  
 OS Prochlorococcus marinus.  
 OC Bacteria; Cyanobacteria; Prochlorophytes; Prochlorococcaceae;  
 CC Prochlorococcus.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CCMP 1375;  
 RX MEDLINE; 98223328.  
 RA RICHTER S., HESS W.R., KRAUSE M., MESSER W.;  
 RT "Unique organization of the dnaA region from Prochlorococcus marinus  
 RT CCMP1375, a marine cyanobacterium."  
 RL Mol. Gen. Genet. 257:534-541(1998).  
 CC -|- SIMILARITY: BELONGS TO THE YCF52 FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; U44977; AAC15819.1; -;  
 DR PFM; PF00583; Acetyltransf; 1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 180 AA; 20564 MW; 3298F75F CRC32;  
 Query Match 100.0%; Score 49; DB 1; Length 180;  
 Best Local Similarity 40.0%; Pred. No. 4.51e+01;

Matches 4; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
 Db 169 PKGNRCAFWY 178  
 QY 2 PXXXXXXFWY 11  
 RESULT 4  
 ID ERS1.YEAST STANDARD; PRT; 260 AA.  
 AC P17261;  
 DT 01-AUG-1990 (Rel. 15, Created)  
 DT 01-AUG-1990 (Rel. 15, Last sequence update)  
 DE TRANSMEMBRANE PROTEIN ERS1 (ERD SUPPRESSOR).  
 DE ERS1 OR YCR075C OR YCR75C.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 CC Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;  
 CC Saccharomycetaceae; Saccharomycetes.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 90245671.  
 RA HARDWICK K., PELHAM H.;  
 RT "ERS1 a seven transmembrane domain protein from Saccharomyces  
 RT cerevisiae."  
 RL Nucleic Acids Res. 18:2177-2177(1990).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA BALLESTA J.P.G., FRANCO L., HOENICKA J., JIMENEZ A., REMACHA M.;  
 RL Submitted (MAR-1992) to the EMBL/GenBank/DBJ databases.  
 CC -|- FUNCTION: SUPPRESSOR OF YEAST ERD1 (REQUIRED FOR THE RETENTION  
 CC OF ENDOGENOUS ER PROTEINS).  
 CC -|- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X52468; CAA36706.1; -;  
 DR EMBL; X59720; CAA42264.1; -;  
 DR PIR; S19490; S19490.  
 DR PIR; S22850; S22850.  
 DR SGD; L0000582; ERS1.  
 KW Transmembrane.  
 FT TRANSMEM 7 28 POTENTIAL.  
 FT TRANSMEM 40 62 POTENTIAL.  
 FT TRANSMEM 81 102 POTENTIAL.  
 FT TRANSMEM 118 138 POTENTIAL.  
 FT TRANSMEM 151 175 POTENTIAL.  
 FT TRANSMEM 185 205 POTENTIAL.  
 FT TRANSMEM 227 247 POTENTIAL.  
 SQ SEQUENCE 260 AA; 30116 MW; E0EA0D6F CRC32;  
 Query Match 100.0%; Score 49; DB 1; Length 260;  
 Best Local Similarity 40.0%; Pred. No. 4.51e+01;  
 Matches 4; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
 Db 74 PKLTQDFWY 83  
 QY 2 PXXXXXXFWY 11  
 RESULT 5  
 ID YL1L.CAEEL STANDARD; PRT; 309 AA.  
 AC Q11098;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE HYPOTHETICAL 35.3 KD PROTEIN C02F12.1 IN CHROMOSOME X.  
 GN C02F12.1.

\*\*\*\*\*

W P E L H  
(TM)

\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Sat Apr 15 01:35:18 2000; MasPar time 3.22 Seconds  
Tabular output not generated. 101.979 Million cell updates/sec

Title: >US-08-452-843-24  
Description: (1-11) from US08452843.pep  
Perfect Score: 49  
Sequence: 1 XPXXXXXXFWY 11

Scoring table: PAM 150  
Gap 15

Searched: 82229 seqs, 29864866 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: swiss-prot38  
1:swissprot

Statistics: Mean 22.187; Variance 40.626; scale 0.546

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	49	100.0	171	1 Y52L_SINY3	YCF52-LIKE PROTEIN.	4.51e+01
2	49	100.0	174	1 YC52_PORPU	HYPOTHETICAL 20.1 KD P	4.51e+01
3	49	100.0	180	1 Y52L_PROMA	YCF52-LIKE PROTEIN.	4.51e+01
4	49	100.0	260	1 ERS1_YEAST	TRANSMEMBRANE PROTEIN	4.51e+01
5	49	100.0	309	1 Y111_CAEEL	HYPOTHETICAL 35.3 KD P	4.51e+01
6	49	100.0	362	1 YHDX_ECOLI	HYPOTHETICAL AMINO-ACI	4.51e+01
7	49	100.0	391	1 KR14_HSV1	GENE 14 PROTEIN KINASE	4.51e+01
8	49	100.0	400	1 RAPQ_RHLY	GENERAL L-AMINO ACID T	4.51e+01
9	49	100.0	426	1 PSGB_HUMAN	PREGNANCY-SPECIFIC BET	4.51e+01
10	49	100.0	430	1 AMPP_HAEIN	XAA-PRO AMINOPEPTIDASE	4.51e+01
11	49	100.0	440	1 AMPP_ECOLI	XAA-PRO AMINOPEPTIDASE	4.51e+01
12	49	100.0	481	1 VGLY_JUNIN	GLYCOPROTEIN POLYPROTE	4.51e+01
13	49	100.0	482	1 VGLY_TACV7	GLYCOPROTEIN POLYPROTE	4.51e+01
14	49	100.0	483	1 VGLY_TACV5	GLYCOPROTEIN POLYPROTE	4.51e+01
15	49	100.0	483	1 VGLY_TACV5	GLYCOPROTEIN POLYPROTE	4.51e+01
16	49	100.0	489	1 VGLY_MOPEI	GLYCOPROTEIN POLYPROTE	4.51e+01
17	49	100.0	495	1 VGLY_TACV	GLYCOPROTEIN POLYPROTE	4.51e+01
18	49	100.0	498	1 VGLY_LYCVW	GLYCOPROTEIN POLYPROTE	4.51e+01
19	49	100.0	498	1 VGLY_LYCVW	GLYCOPROTEIN POLYPROTE	4.51e+01
20	49	100.0	503	1 VGLY_PIARY	GLYCOPROTEIN POLYPROTE	4.51e+01
21	49	100.0	510	1 DCEI_HUMAN	GLUTAMATE DECARBOXYLAS	4.51e+01
22	49	100.0	516	1 P4HA_CHICK	PROLYL 4-HYDROXYLASE A	4.51e+01
23	49	100.0	526	1 P4H1_MOUSE	PROLYL 4-HYDROXYLASE A	4.51e+01

24	49	100.0	527	1 PTIB_BACSU	PTS SYSTEM, ARBUTIN-LI	4.51e+01
25	49	100.0	534	1 P4HA_RAT	PROLYL 4-HYDROXYLASE A	4.51e+01
26	49	100.0	534	1 P4HA_HUMAN	PROLYL 4-HYDROXYLASE A	4.51e+01
27	49	100.0	537	1 P4H2_MOUSE	PROLYL 4-HYDROXYLASE A	4.51e+01
28	49	100.0	538	1 P4HA_CAEEL	PROLYL 4-HYDROXYLASE A	4.51e+01
29	49	100.0	585	1 DCE2_PIG	GLUTAMATE DECARBOXYLAS	4.51e+01
30	49	100.0	585	1 DCE2_HUMAN	GLUTAMATE DECARBOXYLAS	4.51e+01
31	49	100.0	593	1 DCE1_RAT	GLUTAMATE DECARBOXYLAS	4.51e+01
32	49	100.0	593	1 DCE1_MOUSE	GLUTAMATE DECARBOXYLAS	4.51e+01
33	49	100.0	594	1 DCE1_FELCA	GLUTAMATE DECARBOXYLAS	4.51e+01
34	49	100.0	594	1 DCE1_HUMAN	GLUTAMATE DECARBOXYLAS	4.51e+01
35	49	100.0	594	1 DCE1_PIG	GLUTAMATE DECARBOXYLAS	4.51e+01
36	49	100.0	684	1 TC10_YEAST	TCM10 PROTEIN.	4.51e+01
37	49	100.0	733	1 MK10_YEAST	GLUCOSE REPRESSIBLE PR	4.51e+01
38	49	100.0	798	1 VP16_YEAST	VACUOLAR PROTEIN SORTI	4.51e+01
39	49	100.0	1035	1 ENTK_BOVIN	ENTEROPEPTIDASE PRECUR	4.51e+01
40	49	100.0	1748	1 YNR2_YEAST	HYPOTHETICAL 196.1 KD	4.51e+01
41	45	91.8	289	1 SSRL_FUGRU	SOMATOSTATIN-LIKE RECE	1.64e+02
42	45	91.8	448	1 NCAP_CVHOC	NUCLEOCAPSID PROTEIN.	1.64e+02
43	45	91.8	448	1 NCAP_CVTKE	NUCLEOCAPSID PROTEIN.	1.64e+02
44	45	91.8	448	1 NCAP_CVBF	NUCLEOCAPSID PROTEIN.	1.64e+02
45	45	91.8	454	1 NCAP_CVM3	NUCLEOCAPSID PROTEIN.	1.64e+02

ALIGNMENTS

RESULT 1  
ID Y52L\_SINY3 STANDARD; PRT; 171 AA.  
AC Q55911;  
DT 15-JUL-1999 (Rel. 38, Created)  
DT 15-JUL-1999 (Rel. 38, Last sequence update)  
DE YCF52-LIKE PROTEIN.  
GN SLL0286  
OS Synechocystis sp. (strain PCC 6803).  
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.  
RP [1]  
RX MEDLINE: 96127529.  
RA KANEKO T., TANAKA A., SATO S., KOTANI H., SAZUKA T., MIYAJIMA N.,  
RA SUGIURA M., TABATA S.,  
RT 'Sequence analysis of the genome of the unicellular cyanobacterium  
RT Synechocystis sp. strain PCC6803. I. sequence features in the lmb  
RT region from map positions 64 to 92% of the genome.';  
RL DNA Res. 2:153-166(1995).  
CC -I- SIMILARITY: BELONGS TO THE YCF52 FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to licenses@isb-sib.ch).  
CC  
CC EMBL: D64005; BAAL0674.1; .  
CC DR PFAM: PF00583; Acetyltransf; 1.  
CC KW Hypothetical protein.  
CC SQ SEQUENCE 171 AA; 19789 MW; E109E2A4 CRC32;  
Query Match 100.0%; Score 49; DB 1; Length 171;  
Best Local Similarity 40.0%; Pred. No. 4.51e+01;  
Matches 4; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 160 PEGIKGMFWY 169  
QY 2 PXXXXXXFWY 11  
RESULT 2  
ID YC52\_PORPU STANDARD; PRT; 174 AA.  
AC P51192;  
DT 01-OCT-1996 (Rel. 34, Created)

```

##cross-references GB:U04324
GENETICS
#gene GDB:PSG11
##cross-references GDB:128242; OMIM:176398
#map_position 19q13.2-19q13.2
CLASSIFICATION #superfamily carcinoembryonic antigen; carcinoembryonic
antigen precursor amino-terminal homology; immunoglobulin
homology
alternative splicing; glycoprotein
KEYWORDS
FEATURE
1-138 #domain carcinoembryonic antigen precursor
amino-terminal homology #label CEAN\
255-312 #domain immunoglobulin homology #label IMM2
SUMMARY #length 436 #molecular-weight 49494 #checksum 4554

Query Match 100.0%; Score 49; DB 2; Length 436;
Best Local Similarity 40.0%; Pred. No. 9.43e+01;
Matches 4; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 59 PQNLPQYFWY 68
|
|
QY 2 PXXXXXXFWY 11

Search completed: Sat Apr 15 01:35:01 2000
Job time : 20 secs.

```

```

antigen precursor amino-terminal homology; immunoglobulin
homology
glycoprotein; plasma
1-138 #domain carcinoembryonic antigen precursor
amino-terminal homology #label CEAN\
1-34 #domain signal sequence #status predicted #label SIG\
35-426 #product pregnancy-specific beta-1-glycoprotein 7
#status predicted #label MAT\
162-219 #domain immunoglobulin homology #label IMM1\
255-312 #domain immunoglobulin homology #label IMM2\
347-396 #domain immunoglobulin homology #label IMM3\
104,111,199,268,
303,387 #binding_site carbohydrate (Asn) (covalent) #status
predicted
SUMMARY #length 426 #molecular-weight 48272 #checksum 8645
Query Match 100.0%; Score 49; DB 2; Length 426;
Best Local Similarity 40.0%; Pred. No. 9.43e+01;
Matches 4; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
Db 59 PQLPGYFWY 68
|
|
|
QY 2 PXXXXXXFWY 11
|
|
|
RESULT 13 #type complete
ENTRY pregnancy-specific beta-1-glycoprotein 11 form s precursor
TITLE human
ORGANISM #formal_name Homo sapiens #common_name man
DATE 23-Mar-1995 #sequence_revision 23-Mar-1995 #text_change
17-Mar-1999
ACCESSIONS C55181
REFERENCE A55181
#authors McLenachan, P.A.; Rutherford, K.J.; Beggs, K.T.; Sims, S.E.;
Mansfield, B.C.
#journal Genomics (1994) 22:356-363
#title Characterization of the PSG11 gene.
#cross-references MUID:95104846
#accession C55181
#molecule_type DNA
#residues 1-426 #label MCL
#cross-references GB:U04324
REFERENCE S23503
#authors Brophy, B.K.; MacDonald, R.E.; McLenachan, P.A.; Mansfield,
B.C.
#journal Biochim. Biophys. Acta (1992) 1131:119-121
#title cDNA sequence of the pregnancy-specific beta
(1)-glycoprotein-11s (PSG-11s).
#cross-references MUID:92256483
#accession S23503
#molecule_type mRNA
#residues 1-426 #label BRO
#cross-references EMBL:M58591
GENETICS
#gene GDB:PSG11
#cross-references GDB:128242; OMIM:176398
#map_position 19q13.2-19q13.2
CLASSIFICATION #superfamily carcinoembryonic antigen; carcinoembryonic
antigen precursor amino-terminal homology; immunoglobulin
homology
alternative splicing; glycoprotein; plasma
KEYWORDS
FEATURE
1-138 #domain carcinoembryonic antigen precursor
amino-terminal homology #label CEAN\
1-35 #domain signal sequence #status predicted #label SIG\
36-426 #product pregnancy-specific beta-1 glycoprotein 11 form
s #status predicted #label MAT\
162-219 #domain immunoglobulin homology #label IMM1\
255-312 #domain immunoglobulin homology #label IMM2\
347-396 #domain immunoglobulin homology #label IMM3\
104,111,199,268,

```

```

303,387 #binding_site carbohydrate (Asn) (covalent) #status
predicted
SUMMARY #length 426 #molecular-weight 48306 #checksum 8621
Query Match 100.0%; Score 49; DB 2; Length 426;
Best Local Similarity 40.0%; Pred. No. 9.43e+01;
Matches 4; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
Db 59 PQLPGYFWY 68
|
|
|
QY 2 PXXXXXXFWY 11
|
|
|
RESULT 14 #type complete
ENTRY X-pro aminopeptidase (EC 3.4.11.9) II - Haemophilus
TITLE influenzae (strain Rd KW20)
ORGANISM #formal_name Haemophilus influenzae
DATE 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change
10-Oct-1997
ACCESSIONS B64096
REFERENCE A64000
#authors Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.;
Kirkness, E.F.; Kerlavage, A.R.; Sult, C.J.; Tomb, J.F.;
Dougherty, B.A.; Merrick, J.M.; McKenney, K.; Sutton, G.;
FitzHugh, W.; Fields, C.; Gocayne, J.D.; Scott, J.;
Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman,
J.F.; Phillips, C.A.; Spriggs, T.; Hedblom, E.; Cotton,
M.D.; Utterback, T.R.; Hanna, M.C.; Nguyen, D.T.; Saudak,
D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann,
J.L.; Geoghagen, N.S.M.; Gnehm, C.L.; McDonald, L.A.;
Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter, J.C.
#journal Science (1995) 269:496-512
#title Whole-genome random sequencing and assembly of Haemophilus
influenzae Rd.
#cross-references MUID:95350630
#accession B64096
#status nucleic acid sequence not shown; translation not shown
#molecule_type DNA
#residues 1-430 #label TIGR
#cross-references GB:U32764; GB:L42023; NID:G1573827; PID:G1573829;
TIGR:HI0816
CLASSIFICATION #superfamily aminopeptidase P
KEYWORDS alpha-aminoacylpeptide hydrolase; zinc
SUMMARY #length 430 #molecular-weight 49261 #checksum 3043
Query Match 100.0%; Score 49; DB 2; Length 430;
Best Local Similarity 40.0%; Pred. No. 9.43e+01;
Matches 4; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
Db 48 PFRQDSYFWY 57
|
|
|
QY 2 PXXXXXXFWY 11
|
|
|
RESULT 15 #type complete
ENTRY pregnancy-specific beta-1-glycoprotein 11 form r precursor -
TITLE human
ORGANISM #formal_name Homo sapiens #common_name man
DATE 23-Mar-1995 #sequence_revision 23-Mar-1995 #text_change
31-Oct-1997
ACCESSIONS B55181
REFERENCE A55181
#authors McLenachan, P.A.; Rutherford, K.J.; Beggs, K.T.; Sims, S.E.;
Mansfield, B.C.
#journal Genomics (1994) 22:356-363
#title Characterization of the PSG11 gene.
#cross-references MUID:95104846
#accession B55181
#status preliminary
#molecule_type DNA
#residues 1-436 #label MCL

```



```

##experimental_source strain K-12, substrain MG1655
GENETICS
#gene
SUMMARY
    yhdX      #length 368 #molecular-weight 40395 #checksum 8517
    Query Match      100.0%; Score 49; DB 2; Length 368;
    Best Local Similarity 40.0%; Pred. No. 9.43e+01;
    Matches 4; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 141 PPLQIIFWY 150
QY 2 PXXXXXXFWY 11

RESULT 10
ENTRY TVBE11 #type complete
TITLE 44K protein kinase (EC 2.7.1.-) - ictalurid herpesvirus 1
        (strain auburn 1)
ORGANISM #formal_name ictalurid herpesvirus 1
#note host ictalurus punctatus (Channel catfish)
DATE 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change
05-Sep-1997
ACCESSIONS F36787
REFERENCE Davison, A.J.
#authors Davison, A.J.
#submission submitted to GenBank, January 1992
#description Channel catfish virus: a new type of herpesvirus.
#accession F36787
#molecule_type DNA
#residues 1-391 #label DAV
#cross-references GB:M75136; NID:g331209; PID:g331224
REFERENCE A39447
#authors Davison, A.J.
#journal Virology (1992) 186:9-14
#title Channel catfish virus: a new type of herpesvirus.
#cross-references MUID:92087490
#contents annotation
#note neither amino acid nor nucleotide sequence is given
GENETICS
#gene
CLASSIFICATION #superfamily ictalurid herpesvirus 44K protein kinase
KEYWORDS superphototransferase; serine/threonine-specific protein kinase
SUMMARY #length 391 #molecular-weight 44003 #checksum 765

Query Match      100.0%; Score 49; DB 1; Length 391;
Best Local Similarity 40.0%; Pred. No. 9.43e+01;
Matches 4; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 261 PGETMTSFY 270
QY 2 PXXXXXXFWY 11

RESULT 11
ENTRY A54312 #type complete
TITLE pregnancy-specific beta-1 glycoprotein 2 precursor, placental
        (clone hp591) - human
ALTERNATE_NAMES PSG2
ORGANISM #formal_name Homo sapiens #common_name man
DATE 09-Sep-1994 #sequence_revision 09-Sep-1994 #text_change
17-Mar-1999
ACCESSIONS A54312
REFERENCE Chan, W.Y.; Zheng, O.X.; McMahon, J.; Tease, L.A.
#authors Chan, W.Y.; Zheng, O.X.; McMahon, J.; Tease, L.A.
#journal Mol. Cell. Biochem. (1991) 106:161-170
#title Characterization of new members of the pregnancy-specific
        beta1-glycoprotein family.
#cross-references MUID:92017749
#accession A54312
#status preliminary
#molecule_type mRNA
#residues 1-402 #label CHA

##cross-references GB:M94890; GB:M37102; NID:g190567; PID:g190568
CLASSIFICATION #superfamily carcinoembryonic antigen; carcinoembryonic
        antigen precursor amino-terminal homology; immunoglobulin
        homology
        duplication; glycoprotein
KEYWORDS
FEATURE 1-138
    1-34 #domain carcinoembryonic antigen precursor
    35-402 #amino-terminal homology #label CEAN\
        #domain signal sequence #status predicted #label SIG\
        #product pregnancy-specific beta-1 glycoprotein 2
        #status predicted #label MAT\
        162-219 #domain immunoglobulin homology #label IMMI\
        254-303 #domain immunoglobulin homology #label IMM2
SUMMARY #length 402 #molecular-weight 45336 #checksum 4851

Query Match      100.0%; Score 49; DB 2; Length 402;
Best Local Similarity 40.0%; Pred. No. 9.43e+01;
Matches 4; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 59 PQNLPGYFWY 68
QY 2 PXXXXXXFWY 11

RESULT 12
ENTRY B35334 #type complete
TITLE pregnancy-specific beta-1-glycoprotein 7 precursor - human
        pregnancy-specific beta-1 glycoprotein B precursor
ALTERNATE_NAMES #formal_name Homo sapiens #common_name man
ORGANISM 16-Sep-1992 #sequence_revision 16-Sep-1992 #text_change
DATE 24-Sep-1998
ACCESSIONS B35334; B32719; A34621
REFERENCE Streydio, C.; Swillens, S.; Georges, M.; Szpirer, C.;
        Vassart, G.
#authors Streydio, C.; Swillens, S.; Georges, M.; Szpirer, C.;
        Vassart, G.
#journal Genomics (1990) 7:661-662
#cross-references MUID:90353970
#contents erratum
#accession B35334
#molecule_type mRNA
#residues 1-426 #label STR
#note this is a revision to reference A32719
REFERENCE A32719
#authors Streydio, C.; Swillens, S.; Georges, M.; Szpirer, C.;
        Vassart, G.
#journal Genomics (1990) 6:579-592
#title Structure, evolution and chromosomal localization of the
        human pregnancy-specific beta1 glycoprotein gene family.
#cross-references MUID:90256167
#accession B32719
#molecule_type mRNA
#residues 1-70,'Q',72-181,'V',183-185,'R',187-188,'M',190-191,'S',
        193-426 #label ST2
#cross-references GB:M34421
#note this sequence has been revised in reference A35334
REFERENCE A34621
#authors Khan, W.N.; Hammarstrom, S.
#journal Biochem. Biophys. Res. Commun. (1990) 168:214-225
#title Identification of a new carcinoembryonic antigen (CEA) family
        member in human fetal liver - cloning and sequence
        determination of pregnancy-specific glycoprotein 7.
#cross-references MUID:90226362
#accession A34621
#molecule_type mRNA
#residues 1-426 #label KHA
#cross-references GB:M34481; NID:g337666; PID:g337667
COMMENT This protein is found in maternal serum during pregnancy.
COMMENT The amino terminus of the mature protein is not blocked.
GENETICS
#gene
GDB:PSG7
#cross-references GDB:128241; OMIM:176396
#map_position 19q13.2-19q13.2
CLASSIFICATION #superfamily carcinoembryonic antigen; carcinoembryonic
```

```
ENTRY          S22850      #type complete
TITLE          ERS1 protein - yeast (Saccharomyces cerevisiae)
ALTERNATE_NAMES protein YCR075C
ORGANISM       #formal_name Saccharomyces cerevisiae
DATE          23-Apr-1993 #sequence_revision 23-Apr-1993 #text_change
               06-Feb-1998
ACCESSIONS     S22850; S19490
REFERENCE      Hardwick, K.G.; Pelham, H.R.B.
               Nucleic Acids Res. (1990) 18:2177
               ERS1 a seven transmembrane domain protein from Saccharomyces
               cerevisiae.
#cross-references MUID:90245671
#accession       S22850
#molecule_type DNA
#residues       1-260 #label HAR
#cross-references EMBL:X52468; NID:g3687; PID:g3688
REFERENCE      S19486
               Ballesta, J.P.G.; Franco, L.; Hoenicka, J.; Jimenez, A.;
               Remacha, M.; Sanz, E.
               submitted to the Protein Sequence Database, March 1992
               S19490
#molecule_type DNA
#residues       1-260 #label BAL
#cross-references EMBL:X59720; NID:g1907116; PID:e264561; PID:g1907214;
               MIPS:YCR075C
GENETICS       SGD:ERS1
               #cross-references SGD:S0000671; MIPS:YCR075C
#map_position 3R
KEYWORDS       transmembrane protein
FEATURE        41-57
               #domain transmembrane #status predicted #label TM1\
               #domain transmembrane #status predicted #label TM2\
               #domain transmembrane #status predicted #label TM3\
               #domain transmembrane #status predicted #label TM4\
               #domain transmembrane #status predicted #label TM5\
               #domain transmembrane #status predicted #label TM6\
               #length 260 #molecular-weight 30116 #checksum 1512
SUMMARY
Query Match    100.0%; Score 49; DB 2; Length 260;
Best Local Similarity 40.0%; Pred. No. 9.43e+01;
Matches 4; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
Db 74 PKLTQDFWY 83
|
|
|
QY 2 PXXXXXXFWY 11

RESULT 7
ENTRY   JC5050      #type complete
TITLE   sugar phosphate transport protein - Shigella flexneri
ORGANISM #formal_name Shigella flexneri
DATE    31-Jan-1997 #sequence_revision 31-Jan-1997 #text_change
               17-Mar-1999
ACCESSIONS JC5050
REFERENCE   Venkatesan, M.M.; Alexander, W.A.; Fernandez-Prada, C.
               Gene (1996) 175:23-27
               A Shigella flexneri invasion plasmid gene, ipgH, with
               homology to i5629 and sequences encoding bacterial sugar
               phosphate transport proteins.
#cross-references MUID:97074644
#accession       JC5050
#molecule_type DNA
#residues       1-333 #label VEN
#cross-references GB:U28354; NID:g1016674; PID:g1016676
COMMENT        This protein is involved in the uptake of high-energy sugar
               phosphate from an external source.
GENETICS
#gene          ipgH
SUMMARY        #length 333 #molecular-weight 36476 #checksum 3265
```

```
Query Match    100.0%; Score 49; DB 2; Length 333;
Best Local Similarity 40.0%; Pred. No. 9.43e+01;
Matches 4; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
Db 40 PGAKTLVFWY 49
|
|
|
QY 2 PXXXXXXFWY 11

RESULT 8
ENTRY   B34595      #type fragment
TITLE   pregnancy-specific beta-1 glycoprotein 2 - human (fragment)
ORGANISM #formal_name Homo sapiens #common_name man
DATE    07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change
               31-Oct-1997
ACCESSIONS B34595
REFERENCE   A94644
               Zheng, Q.X.; Tease, L.A.; Shupert, W.L.; Chan, W.Y.
               Biochemistry (1990) 29:2845-2852
               Characterization of cDNAs of the human pregnancy-specific
               beta-1-glycoprotein family, a new subfamily of the
               immunoglobulin gene superfamily.
#cross-references MUID:90268037
#accession       B34595
#molecule_type mRNA
#residues       1-351 #label ZHE
#cross-references GB:M31126; NID:g190550; PID:g190551; GB:J02893;
               GB:X15102
CLASSIFICATION #superfamily carcinoembryonic antigen; carcinoembryonic
               antigen precursor amino-terminal homology; immunoglobulin
               homology
               glycoprotein
KEYWORDS       1-92
               #domain carcinoembryonic antigen precursor
               amino-terminal homology (fragment) #label CEAN\
               #domain immunoglobulin homology #label IMM1\
               #domain immunoglobulin homology #label IMM2\
               #binding_site carbohydrate (Asn) (covalent) #status
               predicted
SUMMARY        #length 351 #checksum 8643
Query Match    100.0%; Score 49; DB 2; Length 351;
Best Local Similarity 40.0%; Pred. No. 9.43e+01;
Matches 4; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
Db 8 PQNLPGYFWY 17
|
|
|
QY 2 PXXXXXXFWY 11

RESULT 9
ENTRY   G65119      #type complete
TITLE   hypothetical 40.4 kD protein in acrF-rnd intergenic region -
               Escherichia coli (strain K-12)
ORGANISM #formal_name Escherichia coli
DATE    12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change
               14-Nov-1997
ACCESSIONS G65119
REFERENCE   A84720
               Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.;
               Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.;
               Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.;
               Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao,
               Y.
               Science (1997) 277:1453-1462
               The complete genome sequence of Escherichia coli K-12.
               #cross-references MUID:97426617
               #accession       G65119
               #status        preliminary; nucleic acid sequence not shown;
               translation not shown
               #molecule_type DNA
               #residues       1-368 #label BLAT
               #cross-references GB:AE000405; GB:U000096; NID:g1789659; PID:g1789669;
```

```
translation not shown
##molecule_type DNA
##residues 1-174 ##label REI
##cross-references EMBL:U38804; NID:g1276652; PID:g1276658
##note the nucleotide sequence was submitted to the EMBL Data
Library, October 1995

GENETICS
#genome chloroplast
#keywords chloroplast
#summary #length 174 #molecular-weight 20099 #checksum 7555

Query Match 100.0%; Score 49; DB 2; Length 174;
Best Local Similarity 40.0%; Pred. No. 9.43e+01;
Matches 4; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 163 PDGVKGMFWY 172
Qy 2 PXXXXXXFWY 11

RESULT 3
ENTRY 176669 #type fragment
TITLE pregnancy-specific glycoprotein - mouse (fragment)
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE 02-Aug-1996 #sequence_revision 02-Aug-1996 #text_change
31-Oct-1997
ACCESSIONS 176669
REFERENCE 157007
#authors Rudert, F.; Saunders, A.M.; Thompson, J.A.; Rebstock, S.;
Zimmermann, W.A.
#journal Mamm. Genome (1992) 3:262-273
#title Characterization of murine carcinoembryonic antigen gene
family members.
#cross-references MUID:92345715
#accession 176669
#status preliminary; translated from GB/EMBL/DBDJ
##molecule_type mRNA
##residues 1-209 ##label RES
##cross-references GB:M83346; NID:g200318; PID:g200319
CLASSIFICATION #superfamily carcinoembryonic antigen; carcinoembryonic
antigen precursor amino-terminal homology; immunoglobulin
homology
#keywords glycoprotein
#feature 1-138
#domain carcinoembryonic antigen precursor
amino-terminal homology #label CE1
#summary #length 209 #checksum 5265

Query Match 100.0%; Score 49; DB 2; Length 209;
Best Local Similarity 40.0%; Pred. No. 9.43e+01;
Matches 4; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 180 PKYLSLFWY 189
Qy 2 PXXXXXXFWY 11

RESULT 4
ENTRY 566732 #type complete
TITLE probable membrane protein YOL047c - yeast (Saccharomyces
cerevisiae)
ALTERNATE_NAMES hypothetical protein O2001
ORGANISM #formal_name Saccharomyces cerevisiae
DATE 12-Jul-1996 #sequence_revision 12-Jul-1996 #text_change
14-Nov-1997
ACCESSIONS 566732
REFERENCE 566723
#authors Ansong, W.; Benes, V.; Rechmann, S.; Schwager, C.; Teodoru,
C.; Voss, H.; Wiemann, S.
#submission submitted to the Protein Sequence Database, July 1996
#accession 566732
##molecule_type DNA
##residues 1-234 ##label ANS

##cross-references EMBL:Z74789; NID:g1419849; PID:e251859; PID:g1419850;
MIPS:YOL047c
##experimental_source strain S288C

GENETICS
#map_position 15L
#introns 1/1
#keywords transmembrane protein
#feature 7-23
#domain transmembrane #status predicted #label TM1\
100-116 #domain transmembrane #status predicted #label TM2\
165-181 #domain transmembrane #status predicted #label TM3\
#length 234 #molecular-weight 26460 #checksum 829

SUMMARY
Query Match 100.0%; Score 49; DB 2; Length 234;
Best Local Similarity 40.0%; Pred. No. 9.43e+01;
Matches 4; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 90 PQFFIFFFWY 99
Qy 2 PXXXXXXFWY 11

RESULT 5
ENTRY JC4121 #type complete
TITLE pregnancy-specific glycoprotein 11s' precursor - human
ORGANISM #formal_name Homo sapiens #common_name man
DATE 26-Jul-1995 #sequence_revision 19-Oct-1995 #text_change
17-Mar-1999
ACCESSIONS JC4121
REFERENCE JC4121
#authors Reglund, S.; Zhou, G.O.; Hammarstrom, S.
#journal Biochem. Biophys. Res. Commun. (1995) 211:656-664
#title Characterization of cDNA encoding novel pregnancy-specific
glycoprotein variants.
#cross-references MUID:95314639
#accession JC4121
##molecule_type mRNA
##residues 1-240 ##label TEG
##cross-references GB:U25987; NID:g862680; PID:g862681
#experimental_source fetal liver
#comment This protein belongs to the carcinoembryonic antigen family. This
protein is a molecule synthesized by the placental
syncytiotrophoblasts and released to the maternal circulation
during pregnancy.

GENETICS
#gene GDB:PSG11
#cross-references GDB:128242; OMIM:176398
#map_position 19q13.2-19q13.2
CLASSIFICATION #superfamily carcinoembryonic antigen precursor
amino-terminal homology; immunoglobulin homology
#keywords glycoprotein
#feature 1-138
#domain carcinoembryonic antigen precursor
amino-terminal homology #label CEAN\
#domain signal sequence #status predicted #label SIG\
#product pregnancy-specific glycoprotein 11s' #status
predicted #label MAT\
127-129 #region cell attachment (R-G-D) motif\
161-210 #domain immunoglobulin homology #label IMM\
104,111,201 #binding_site carbohydrate (Asn) (covalent) #status
predicted
SUMMARY #length 240 #molecular-weight 27004 #checksum 9572

Query Match 100.0%; Score 49; DB 2; Length 240;
Best Local Similarity 40.0%; Pred. No. 9.43e+01;
Matches 4; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 59 PQNLPGYFWY 68
Qy 2 PXXXXXXFWY 11

RESULT 6
```

\*\*\*\*\*  
M P S R E H  
\*\*\*\*\*  
(TM)  
\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd  
MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Sat Apr 15 01:34:41 2000; MasPar time 3.22 Seconds  
136.845 Million cell updates/sec  
Tabular output not generated.

Title: >US-08-452-843-24  
Description: (1-11) from US08452843.pep  
Perfect Score: 49  
Sequence: 1 XPXXXXXXFWY 11  
Scoring table: PAM 150  
Gap 15  
Searched: 122810 seqs, 40068593 residues  
Post-processing: Minimum Match 0%  
Listing first 45 summaries  
Database: pir62  
1:pir1 2:pir2 3:pir3 4:pir4  
Statistics: Mean 21.740; Variance 44.150; scale 0.492

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES						
Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	49	100.0	133	2	mp41 - mouse	9.43e+01
2	49	100.0	174	2	hypothetical protein	9.43e+01
3	49	100.0	209	2	pregnancy-specific gl	9.43e+01
4	49	100.0	234	2	probable membrane pro	9.43e+01
5	49	100.0	240	2	pregnancy-specific gl	9.43e+01
6	49	100.0	260	2	ERS1 protein - yeast	9.43e+01
7	49	100.0	333	2	sugar phosphate trans	9.43e+01
8	49	100.0	351	2	pregnancy-specific be	9.43e+01
9	49	100.0	368	2	hypothetical 40.4 kD	9.43e+01
10	49	100.0	391	1	TVBE11	9.43e+01
11	49	100.0	402	2	pregnancy-specific be	9.43e+01
12	49	100.0	426	2	B35334	9.43e+01
13	49	100.0	426	2	C35181	9.43e+01
14	49	100.0	430	2	B64096	9.43e+01
15	49	100.0	436	2	B55181	9.43e+01
16	49	100.0	441	1	DPBPC	9.43e+01
17	49	100.0	482	1	VGXPT7	9.43e+01
18	49	100.0	483	1	VGXPT5	9.43e+01
19	49	100.0	489	1	VGXPMV	9.43e+01
20	49	100.0	495	2	A55181	9.43e+01
21	49	100.0	498	1	VGXPLA	9.43e+01
22	49	100.0	498	1	VGXPLM	9.43e+01
23	49	100.0	503	1	QOXGPP	9.43e+01

24 49 100.0 510 1 A30999 glutamate decarboxyla 9.43e+01  
25 49 100.0 516 1 DACHA procollagen-proline d 9.43e+01  
26 49 100.0 526 2 I49134 prol 4-hydroxylase 9.43e+01  
27 49 100.0 534 1 DAHUA2 procollagen-proline d 9.43e+01  
28 49 100.0 534 1 DAHUA1 procollagen-proline d 9.43e+01  
29 49 100.0 534 2 S44204 procollagen-proline d 9.43e+01  
30 49 100.0 537 2 I49135 prol 4-hydroxylase 9.43e+01  
31 49 100.0 558 2 A55069 procollagen-proline d 9.43e+01  
32 49 100.0 575 1 JH0827 glutamate decarboxyla 9.43e+01  
33 49 100.0 585 1 A41292 glutamate decarboxyla 9.43e+01  
34 49 100.0 585 2 S61534 glutamate decarboxyla 9.43e+01  
35 49 100.0 595 2 JC4064 glutamate decarboxyla 9.43e+01  
36 49 100.0 593 1 A41367 glutamate decarboxyla 9.43e+01  
37 49 100.0 593 2 S48135 glutamate decarboxyla 9.43e+01  
38 49 100.0 594 1 B41935 glutamate decarboxyla 9.43e+01  
39 49 100.0 594 2 S51775 glutamate decarboxyla 9.43e+01  
40 49 100.0 594 1 A46758 glutamate decarboxyla 9.43e+01  
41 49 100.0 611 2 S61147 TCM10 protein - yeast 9.43e+01  
42 49 100.0 733 2 S31288 MAK10 protein - yeast 9.43e+01  
43 49 100.0 798 2 S62031 vacuolar protein sort 9.43e+01  
44 49 100.0 1035 2 A43090 enteropeptidase (EC 3 9.43e+01  
45 49 100.0 1748 2 S63127 probable membrane pro 9.43e+01

ALIGNMENTS

RESULT 1  
ENTRY I52649 #type complete  
TITLE mp41 - mouse  
ORGANISM #formal\_name Mus sp. #common\_name mouse  
DATE 26-Jul-1996 #sequence\_revision 26-Jul-1996 #text\_change 28-Feb-1997  
ACCESSIONS I52649  
REFERENCE I52649  
#authors Ishida, N.; Matsui, M.; Nishimatsu, S.; Murakami, K.; Mitsui, Y.  
#journal Brain Res. Mol. Brain Res. (1994) 26:197-206  
#title Molecular cloning of a gene under control of the circadian clock and light in the rodent SCN.  
#cross-references MUID:95157174  
#accession I52649  
#status preliminary; translated from GB/EMBL/DBJ  
#molecule\_type mrna  
#residues 1-133 #label RES  
#cross-references GB:S76755; NID:g913851; PID:g913852  
GENETICS  
#gene mp41  
SUMMARY #length 133 #molecular-weight 14391 #checksum 258  
Query Match 100.0%; Score 49; DB 2; Length 133;  
Best Local Similarity 40.0%; Pred. No. 9.43e+01;  
Matches 4; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 11 PTTGRIFWY 20  
QY 2 PXXXXXXFWY 11  
RESULT 2  
ENTRY S73113 #type complete  
TITLE hypothetical protein 174 - red alga (Porphyra purpurea) Chloroplast  
ORGANISM #formal\_name chloroplast Porphyra purpurea  
DATE 19-Mar-1997 #sequence\_revision 09-May-1997 #text\_change 10-Sep-1997  
ACCESSIONS S73113  
REFERENCE S73108  
#authors Reith, M.; Munholland, J.  
#journal Plant Mol. Biol. Rep. (1995) 13:333-335  
#title Complete nucleotide sequence of the Porphyra purpurea chloroplast genome.  
#accession S73113  
#status preliminary; nucleic acid sequence not shown;

THIS PAGE BLANK (USPTO)

SQ Sequence 539 AA;

Query Match 100.0%; Score 49; DB 1; Length 539;  
Best Local Similarity 40.0%; Pred. No. 3.50e+02;  
Matches 4; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 117 PAGHAGTFWY 126  
|  
|  
|  
QY 2 PXXXXXXFWY 11

Search completed: Sat Apr 15 01:34:22 2000  
Job time : 36 secs.

Matches 4; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 117 PAGHAGTFWY 126  
|  
|  
|  
QY 2 PXXXXXXFWY 11

## RESULT 14

ID W1973 standard; Protein; 539 AA.  
AC W1973;  
DT 21-JUL-1997 (first entry)  
DE Coprinus cinereus lcc1 polypeptide.  
KW Benzene:O<sub>2</sub> oxygen oxidoreductase; laccase; lignin; Kraft pulp; dye;  
OS Coprinus cinereus chain reaction; papermaking.  
PN WO9708325-A2.  
PD 06-MAR-1997.  
PF 20-AUG-1996; U13728.  
PR 25-AUG-1995; US-002800.  
PA (NOVO ) NOVO NORDISK BIOTECH INC.  
PI Brown KM, Halkier T, Kauppinen S, Yaver DS;  
DR WPI: 97-179282/16.  
DR N-PSDB; T69936.

New laccase from Coprinus strains - useful for polymerising lignin,  
FT depolymerising Kraft pulp, oxidising dyes and their precursors, etc.  
PS Claim 6; Fig 1; 62pp; English.  
CC The present sequence represents a novel laccase, lcc1, isolated from  
CC Coprinus cinereus strain IFO 8371. This polypeptide is used  
CC to polymerise a lignin or lignosulphate in solution; for in situ  
CC depolymerisation of Kraft pulp; for oxidising dyes or their precursors  
CC (particularly to prevent dye transfer between fabrics and in hair dyeing).  
CC and for polymerising or oxidising phenolic compounds (e.g. to  
CC precipitate phenolics from fruit juices to give a more stable product).  
CC It can also be used for soil detoxification. Use of the polypeptide  
CC avoids the need to use chlorine for lignin depolymerisation. It has  
CC better activity than known laccases under the alkaline conditions  
CC usually encountered in papermaking processes. A cDNA library from  
CC IFO 8371 was prepared and subjected to PCR with oligonucleotides  
CC based on the conserved motifs in other fungal laccases. The  
CC amplification product was cloned and 7 subclones were produced and  
CC sequenced. They correspond to 3 different laccases designated lcc1, 2  
CC and 3. To isolate full-length DNA, a genomic DNA library of IFO 8371 was  
CC constructed. A digoxigenin-labelled probe was prepared by PCR using lcc1  
CC cDNA as a template and 32p-labelled probes from lcc2 and 3 partial cDNA.  
CC These probes were used to screen the genomic library and two clones  
CC were isolated, one containing the lcc1 gene and the other containing the  
CC lcc3 gene. No single clone contained the complete lcc2 gene which was  
CC isolated from two partial clones.  
CC N.B. The sequence presented in this record is the same as the  
CC version supplied electronically to the European Patent Office; it  
CC differs from the sequence printed in Figure 1 of the specification.  
SQ Sequence 539 AA;

Query Match 100.0%; Score 49; DB 1; Length 539;

Best Local Similarity 40.0%; Pred. No. 3.60e+02;

Matches 4; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 117 PAGHAGTFWY 126  
|  
|  
|  
QY 2 PXXXXXXFWY 11

## RESULT 15

ID W76282 standard; protein; 539 AA.  
AC W76282;  
DT 08-JAN-1999 (first entry)  
DE Coprinus cinereus laccase protein variant.  
KW Laccase; variant; oxidation; dye transfer inhibition; bleaching;  
KW denim; lignin modification; paper strengthening; phenol polymerisation;  
KW hair dye; waste water treatment.  
OS Synthetic.  
OS Coprinus cinereus.

FT	Key	Location/Qualifiers
FT	Misc_difference 98	/label= D98X
FT		/note= "X is optionally Ala, Val, Pro, Leu, Ile,
FT		Phe, Tyr or Trp"
FT	Misc_difference 131	/label= G131X
FT		/note= "X is optionally Ala, Val, Pro, Leu, Ile,
FT		Phe, Tyr or Trp"
FT	Misc_difference 257	/label= T257X
FT		/note= "X is optionally Ala, Val, Pro, Leu, Ile,
FT		Phe, Tyr or Trp"
FT	Misc_difference 260	/label= R260X
FT		/note= "X is optionally Ala, Val, Pro, Leu, Ile,
FT		Phe, Tyr or Trp"
FT	Misc_difference 351	/label= L351X
FT		/note= "X is optionally Ile, Phe, Tyr or Trp"
FT	Misc_difference 358	/label= F358X
FT		/note= "X is optionally Ile or Trp"
FT	Misc_difference 359	/label= T359
FT		/note= "X is optionally Ala, Val, Pro, Leu, Ile,
FT		Phe, Tyr or Trp"
FT	Misc_difference 409	/label= V409X
FT		/note= "X is optionally Pro, Leu, Ile, Phe, Tyr or Trp"
FT	Misc_difference 411	/label= G411X
FT		/note= "X is optionally Ala, Val, Pro, Leu, Ile,
FT		Phe, Tyr or Trp"
FT	Misc_difference 412	/label= G412X
FT		/note= "X is optionally Ala, Val, Pro, Leu, Ile,
FT		Phe, Tyr or Trp"
FT	Misc_difference 443	/label= D443X
FT		/note= "X is optionally Ala, Val, Pro, Leu, Ile,
FT		Phe, Tyr or Trp"
FT	Misc_difference 473	/label= E473X
FT		/note= "X is optionally Ala, Val, Pro, Leu, Ile,
FT		Phe, Tyr or Trp"
FT	Misc_difference 480	/label= L480X
FT		/note= "X is optionally Phe, Tyr or Trp"
FT	WO9838287-Al.	
PN	03-SEP-1998.	
PD	23-FEB-1998; DK0070.	
PF	28-FEB-1997; DK-000222.	
PR	(NOVO ) NOVO-NORDISK AS.	
PA	Svendsen A, Xu F;	
PI	WPI: 98-495393/42.	
DR	New variants of Coprinus and related laccases with increased	
PT	oxidation potential - or altered pH optimum, or mediator or	
PT	oxygen-hydroxide ion pathways, useful for oxidation, for inhibiting	
PT	dye transfer and in bleaching textiles, especially as detergent	
PT	additive	
PS	Claim 2; Page -; 147pp; English.	
CC	The present sequence represents a mutant laccase protein. The	
CC	specification describes active laccase variants (see W76282,	
CC	W76296-99 and W76316-17) having increased oxidation potential,	
CC	altered pH optimum, altered mediator and/or altered oxygen/hydroxide	
CC	ion pathway. The laccase variants are used specifically to oxidise	
CC	substrates, to inhibit dye transfer, and for bleaching textiles,	
CC	specifically denim. They can also be used for lignin modification,	
CC	strengthening paper, polymerisation of phenols, dyeing of hair and	
CC	textiles and waste water treatment.	
CC	note: the present sequence does not appear in the specification; it was	
CC	constructed using information provided.	





abnormal keratinocyte differentiation; psoriasis; epithelial cancer;  
Parkinson's disease; Alzheimer's disease; ALS; neuropathy;  
fibromodulin; dermal scarring; Usher Syndrome; Atrophia areata;  
anti-thrombotic; wound healing; tissue repair.

ulceration and congenital microvillus atrophy), skin diseases associated  
with abnormal keratinocyte differentiation (e.g. psoriasis, epithelial  
cancers such as lung squamous cell carcinoma of the vulva and gliomas),  
potent effects on cell growth and development, diseases related to growth  
or survival of nerve cells including Parkinson's disease, Alzheimer's  
disease, ALS, neuropathies or cancer. PRO265 can be used as for  
fibromodulin, e.g. for reducing dermal scarring. PRO264 can be used  
as a target for anti-tumor drugs. PRO533 may be used in the treatment  
of Usher Syndrome or Atrophia areata; PRO269 can be used as an  
anti-thrombotic agent; PRO287 polypeptides and portions may have  
therapeutic applications in wound healing and tissue repair; PRO317 can  
be used for treating problems of the kidney, uterus, endometrium, blood  
vessels, or related tissue, e.g. in the heart of genital tract.

Query Match 100.0%; Score 49; DB 1; Length 533;  
Best Local Similarity 40.0%; Pred. No. 3.60e+02;  
Matches 4; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 476 PKKGTAVFWY 485  
|  
|  
|  
Qy 2 PXXXXXXFWY 11

RESULT 10  
ID W37046 standard; Protein; 535 AA.  
AC W37046;  
DT 28-MAY-1998 (first entry)  
DE Human alpha-(2) subunit of propyl-4-hydrolase.  
KW Alpha-(2) subunit; propyl-4-hydrolase; collagen synthesis;  
KW 4-hydroxyproline; human.  
OS Homo sapiens.

FS Key Location/Qualifiers  
FT CDS 187..1795  
FT /\*tag= a  
FT /product= Alpha-(2) propyl-4-hydrolase

FN W09738121-A1.  
PD 16-OCT-1997.  
PF 18-MAR-1997; U04358.  
PR 10-APR-1996; US-633879.  
PA (FIBR-) FIBROGEN INC.  
PI Annunen PP, Helakoski TI, Kivirikko KI, Nissi RK,  
PI Nokelainen MK, Pihlajaniemi TA;  
DR WPI; 97-326117/48.  
DR N-PSDB; V00500.  
PT Human prolyl-4-hydroxylase alpha-2 subunit - used for the  
production of correctly folded collagen  
PS Claim 3; Pages 36-40; 85pp; English.  
CC This is a human alpha-(2) subunit of propyl-4-hydrolase. The  
alpha-(2) subunit of propyl-4-hydrolase is part of an enzyme, which  
plays a crucial role in the synthesis of all collagens. Specifically,  
the enzyme catalyses the formation of 4-hydroxyproline in collagens,  
which is essential for the folding of newly synthesised collagen  
polypeptide chains into triple-helical molecules.

Query Match 100.0%; Score 49; DB 1; Length 535;  
Best Local Similarity 40.0%; Pred. No. 3.60e+02;  
Matches 4; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 478 PKKGTAVFWY 487  
|  
|  
|  
Qy 2 PXXXXXXFWY 11

RESULT 11  
ID W37045 standard; Protein; 537 AA.  
AC W37045;  
DT 28-MAY-1998 (first entry)  
DE Murine alpha-(2) subunit of propyl-4-hydrolase.  
KW Alpha-(2) subunit; propyl-4-hydrolase; collagen synthesis;  
KW 4-hydroxyproline; mouse.  
OS Murine.

abnormal keratinocyte differentiation; psoriasis; epithelial cancer;  
Parkinson's disease; Alzheimer's disease; ALS; neuropathy;  
fibromodulin; dermal scarring; Usher Syndrome; Atrophia areata;  
anti-thrombotic; wound healing; tissue repair.

abnormal keratinocyte differentiation; psoriasis; epithelial cancer;  
Parkinson's disease; Alzheimer's disease; ALS; neuropathy;  
fibromodulin; dermal scarring; Usher Syndrome; Atrophia areata;  
anti-thrombotic; wound healing; tissue repair.

ulceration and congenital microvillus atrophy), skin diseases associated  
with abnormal keratinocyte differentiation (e.g. psoriasis, epithelial  
cancers such as lung squamous cell carcinoma of the vulva and gliomas),  
potent effects on cell growth and development, diseases related to growth  
or survival of nerve cells including Parkinson's disease, Alzheimer's  
disease, ALS, neuropathies or cancer. PRO265 can be used as for  
fibromodulin, e.g. for reducing dermal scarring. PRO264 can be used  
as a target for anti-tumor drugs. PRO533 may be used in the treatment  
of Usher Syndrome or Atrophia areata; PRO269 can be used as an  
anti-thrombotic agent; PRO287 polypeptides and portions may have  
therapeutic applications in wound healing and tissue repair; PRO317 can  
be used for treating problems of the kidney, uterus, endometrium, blood  
vessels, or related tissue, e.g. in the heart of genital tract.

CC of tumours. The protein has immunosuppressive activity and growth  
 CC promoting activity and can be used for enhancing fertility in  
 CC females, viability of a foetus etc.  
 CC See aslo R06428-34.  
 SQ Sequence 354 AA;

Query Match 100.0%; Score 49; DB 1; Length 354;  
 Best Local Similarity 40.0%; Pred. No. 3.60e+02;  
 Matches 4; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 11 PQNLPGVFWY 20  
 |  
 |||  
 QY 2 PXXXXXXFWY 11

## RESULT 6

ID R53700 standard; Protein; 387 AA.

AC R53700;  
 DT 09-NOV-1994 (first entry)  
 DE Sequence of corn microsomal delta-12 desaturase deduced from the  
 DE cDNA in plasmid pRad2 1.  
 KW Fatty acid; desaturase; lipid; unsaturated; transgenic plant.  
 OS Zea mays.  
 PN W09411516-A.  
 PD 26-MAY-1994.  
 PF 15-OCT-1993; U09987.  
 PR 17-NOV-1992; US-977339.  
 PA (DUPO ) DU PONT DE NEMOURS & CO E I.  
 PI Lightner JE, Okuley JJ;  
 DR WPI; 94-183515/22.  
 DR N-PSDB; Q66071.

PT Genes for fatty acid desaturase enzymes - permit alteration of  
 PT plant lipid composition

PS Claim 13; Page 124-126; 147pp; English.  
 CC Corn microsomal delta-12 desaturase cDNA was isolated using a PCR  
 CC approach. A cDNA library was made to poly A+ mRNA from developing  
 CC corn embryos. This library was used as template for PCR using sets  
 CC of degenerate oligos NS3 (Q66075) and RB5A/B (Q66077, Q66078) as  
 CC sense and antisense primers, respectively. NS3 and RB5A/B corresp.  
 CC to stretches of AAs 101-109 and 318-326, respectively, of R53697.  
 CC which are conserved in most microsomal delta-12 desaturases. A PCR  
 CC product of 720bp was purified and used as a probe for screening the  
 CC corn cDNA library. A plaque was purified and found to encode  
 CC microsomal delta-12 desaturase truncated at the 3' end. This cDNA  
 CC was used to probe the corn cDNA library again. The clone contg.  
 CC the longest insert, designated pRad2 1 was sequenced completely  
 CC (Q66071). An isolated nucleic acid fragment  
 CC wherein the nucleic acid identity is 90% or greater to  
 CC Q66071 is claimed. A method to isolate nucleic acid fragments  
 CC encoding fatty acid desaturases and related enzymes is claimed  
 CC which comprises: comparing AA sequences in R53697-R53702 and other  
 CC fatty acid desaturase sequences; identifying conserved sequences of  
 CC 4 or more AAs; designed degenerate oligos based on the conserved  
 CC sequences; and using the oligos to isolated sequences encoding fatty  
 CC acid desaturases and desaturase-related enzymes.  
 SQ Sequence 387 AA;

Query Match 100.0%; Score 49; DB 1; Length 387;  
 Best Local Similarity 40.0%; Pred. No. 3.60e+02;  
 Matches 4; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 374 PEDRKGVFWY 383  
 |  
 |||  
 QY 2 PXXXXXXFWY 11

## RESULT 7

ID R05272 standard; protein; 441 AA.

AC R05272;  
 DT 15-AUG-1990 (first entry)  
 DE Polypeptide with amino peptidase-P activity encoded by new gene  
 KW Amino peptidase-P.  
 PN J02002373-A.

PD 08-JAN-1990.  
 PF 25-MAR-1989; 071138.  
 PR 25-MAR-1989; JP-071138, JP-156193.  
 PA (AJIN) Ajinomoto Kk.  
 PI WPI; 90-053424/08.  
 DR N-PSDB; Q91838.  
 PT Amino peptidase-P-coding gene -  
 PT used in gene-provided recombinant DNA and recombinant  
 PT DNA-provided survival cell stock  
 PS Disclosure; 15pp; Japanese.  
 CC It is new. Also new are recombinant DNA contg. its encoding DNA, cells  
 CC transformed with the recombinant DNA, and prodn. of it by culturing the  
 CC cells. The method allows economical, high yielding prodn. of it. It is  
 CC also useful in separating or refining the enzyme.  
 SQ Sequence 441 AA;

Query Match 100.0%; Score 49; DB 1; Length 441;  
 Best Local Similarity 40.0%; Pred. No. 3.60e+02;  
 Matches 4; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 43 PYRQNSDFWY 52  
 |  
 |||  
 QY 2 PXXXXXXFWY 11

## RESULT 8

ID W51087 standard; Protein; 451 AA.

AC W51087;  
 DT 27-AUG-1998 (first entry)  
 DE Delta 9 desaturase amino acid sequence.  
 KW Delta 9 desaturase; higher substrate specificity; stearic acid;  
 KW palmitic acid; recombinant yeast cell; high resistance; low temperature;  
 KW fermentation process; fatty acid composition; high level; oleic acid;  
 KW palmitoleic acid.  
 OS Pichia angusta.  
 PN J10075782-A.  
 PD 24-MAR-1998.  
 PF 04-SEP-1996; 270405.  
 PR 04-SEP-1996; JP-270405.  
 PA (SHOS ) SHOMA SANGYO CO.  
 DR WPI; 98-244358/22.  
 DR N-PSDB; V07175.

PT New Pichia angusta delta-9 desaturase gene - confers resistance to  
 PT very low temperatures; used to produce fatty acid compositions  
 PT containing more oleic acid than palmitic acid

PS Claim 2; Pages 12-14; 15pp; Japanese.  
 CC The present sequence represents a new Delta 9 desaturase gene of Pichia  
 CC angusta. This Delta 9 desaturase that has a higher substrate specificity  
 CC to stearic acid than palmitic acid. The Delta 9 desaturase gene is used  
 CC to produce recombinant yeast cells that have a high resistance to low  
 CC temperatures. These yeast cells can be used in a fermentation process to  
 CC produce fatty acid compositions that contain a higher level of oleic acid  
 CC than palmitoleic acid.  
 SQ Sequence 451 AA;

Query Match 100.0%; Score 49; DB 1; Length 451;  
 Best Local Similarity 40.0%; Pred. No. 3.60e+02;  
 Matches 4; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 151 PYDAKRGFWY 160  
 |  
 |||  
 QY 2 PXXXXXXFWY 11

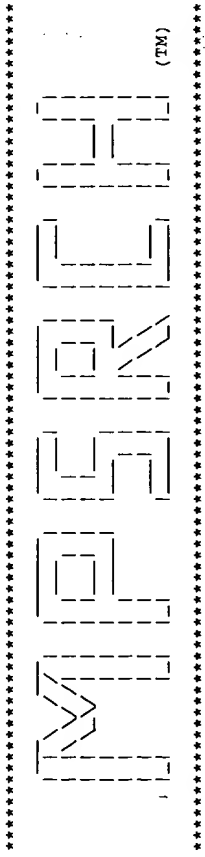
## RESULT 9

ID Y13400 standard; Protein; 533 AA.

AC Y13400;  
 DT 25-JUN-1999 (first entry)  
 DE Amino acid sequence of protein PRO330.  
 KW Secreted protein; transmembrane protein; human; enterocolitis;  
 KW Zollinger-Ellison syndrome; gastrointestinal ulceration;  
 KW congenital microvillus atrophy; skin disease; cell growth;

RESULT	4
ID	R59522 standard; protein; 341 AA.
AC	R59522;
DE	09-Nov-1994 (first entry)
DT	GAD65 1-244 N-terminal-deleted mutant.
DE	GAD65; glutamate-decarboxylase; diabetes mellitus;
KW	stiff man syndrome; autoantibody; mutagenesis.
OS	Homo sapiens.
PN	WO9412529-A.

CC (SP1) also known as pregnancy-specific beta glycoprotein (PSBG),  
CC detected in placenta and in testis. The sequence and Abs specific  
CC for it can be used in diagnosis, pregnancy testing and monitoring  
CC



\*\*\*\*\*  
Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd  
\*\*\*\*\*  
MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Sat Apr 15 01:33:46 2000; MasPar time 3.09 Seconds  
Tabular output not generated. 84.386 Million cell updates/sec

Title: >US-08-452-843-24  
Description: (1-11) from US08452843.pep  
Perfect Score: 49  
Sequence: 1 XPXXXXXFWY 11  
  
Scoring table: PAM 150  
Gap 15  
  
Searched: 188963 seqs, 23686106 residues  
Post-processing: Minimum Match 0%  
Listing first 45 summaries  
  
Database: a-geneseq36  
1:geneseqp  
  
Statistics: Mean 15.726; Variance 64.556; scale 0.244

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	49	100.0	20	1	Glutamic acid decarboxylase	3.60e+02
2	49	100.0	87	1	Streptococcus pneumoniae	3.60e+02
3	49	100.0	133	1	Anticancer polypeptide	3.60e+02
4	49	100.0	341	1	GAD65 1-244 N-terminal	3.60e+02
5	49	100.0	354	1	SPI-like protein encod	3.60e+02
6	49	100.0	387	1	Sequence of corn micro	3.60e+02
7	49	100.0	441	1	Polypeptide with amino	3.60e+02
8	49	100.0	451	1	Delta 9 desaturase ami	3.60e+02
9	49	100.0	533	1	Amino acid sequence of	3.60e+02
10	49	100.0	535	1	Human alpha-(2) subun	3.60e+02
11	49	100.0	537	1	Murine alpha-(2) subun	3.60e+02
12	49	100.0	539	1	Coprinus cinereus lacc	3.60e+02
13	49	100.0	539	1	Coprinus cinereus lacc	3.60e+02
14	49	100.0	539	1	Coprinus cinereus lcc1	3.60e+02
15	49	100.0	539	1	Coprinus cinereus lacc	3.60e+02
16	49	100.0	540	1	GAD65 1-45 N-terminal	3.60e+02
17	49	100.0	544	1	GAD65 545-585 C-termin	3.60e+02
18	49	100.0	554	1	GAD65 1-31 Deleted, C4	3.60e+02
19	49	100.0	584	1	Human GAD65 protein.	3.60e+02
20	49	100.0	584	1	Human GAD65 protein se	3.60e+02
21	49	100.0	585	1	Human GAD 65	3.60e+02
22	49	100.0	585	1	Modified glutamic acid	3.60e+02
23	49	100.0	585	1	Human 65K-glutamic aci	3.60e+02

24	49	100.0	585	1	R28756	Human pancreatic islet	3.60e+02
25	49	100.0	585	1	R71541	Human GAD.	3.60e+02
26	49	100.0	585	1	W12402	65 kD human glutamic a	3.60e+02
27	49	100.0	585	1	R59516	Human GAD65.	3.60e+02
28	49	100.0	585	1	W14916	Modified glutamic acid	3.60e+02
29	49	100.0	585	1	R79105	Human glutamic acid de	3.60e+02
30	49	100.0	593	1	R27220	Brain GAD #2.	3.60e+02
31	49	100.0	594	1	W74717	Amino acid sequence of	3.60e+02
32	49	100.0	594	1	R27221	Full length brain GAD.	3.60e+02
33	49	100.0	594	1	R27222	Full length islet GAD.	3.60e+02
34	49	100.0	788	1	R57283	Bovine enterokinase.	3.60e+02
35	45	91.8	65	1	W96258	Human semaphorin recep	7.66e+02
36	45	91.8	230	1	R47151	IL-2 receptor gamma ch	7.66e+02
37	45	91.8	402	1	R63142	Glycoprotein 50 (gp50)	7.66e+02
38	45	91.8	402	1	P70644	Pseudorabies virus gp5	7.66e+02
39	45	91.8	475	1	W97590	Protein encoded by a f	7.66e+02
40	45	91.8	475	1	W97589	Full length Cry6A prot	7.66e+02
41	45	91.8	475	1	R26226	Delta endotoxin.	7.66e+02
42	45	91.8	901	1	W96256	Mouse semaphorin recep	7.66e+02
43	45	91.8	906	1	W96257	Mouse semaphorin recep	7.66e+02
44	45	91.8	1194	1	W91071	Apoptosis inducer Apaf	7.66e+02
45	45	91.8	1205	1	W91072	Apoptosis inducer spl1	7.66e+02

ALIGNMENTS

RESULT 1  
ID R72293 standard; Peptide; 20 AA.  
AC R72293;  
DT 13-NOV-1995 (first entry)  
DE Glutamic acid decarboxylase (GAD65) fragment.  
KW Glutamic acid decarboxylase; GAD65; autoimmune disorders;  
KW Insulin-dependent diabetes mellitus; stiff man disease.  
OS Homo sapiens.  
PN W09507992-A.  
PD 23-MAR-1995.  
PF 24-AUG-1994; U09478.  
PR 17-SEP-1993; US-123859.  
PA (REGC ) UNIV CALIFORNIA.  
PI Clare-Salzler MJ, Erlander MG, Kaufman DL, Tobin AJ;  
WPI: 95-131360/17.  
DR New polypeptide fragments of glutamic acid decarboxylase - for  
PT diagnosis and treatment of autoimmune disease, esp. insulin  
PT dependent diabetes, also related nucleic acid, vectors,  
PT antibodies, hybridoma(s) etc.  
PS Example 11; Page 76; 100pp; English.  
CC O86481 and O86482 encode R71733 and R79105, rat and human glutamic  
CC acid decarboxylase (GAD65), respectively, from which the GAD65  
CC fragments described in R72261-R72298 were derived. These fragments  
CC can be used to detect autoantibodies against GAD, e.g. to diagnose  
CC and treat GAD-related autoimmune disorders, such as insulin  
CC dependant diabetes mellitus or stiff man disease.  
SQ Sequence 20 AA;

Query Match 100.0%; Score 49; DB 1; Length 20;  
Best Local Similarity 40.0%; Pred. No. 3.60e+02;  
Matches 4; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
Db 6 PQHTNCFWY 15  
QY 2 PXXXXXFWY 11

RESULT	2	
ID	Y11265	standard; Protein; 87 AA.
AC	Y11265;	
DT	20-MAY-1999	(first entry)
DE	Streptococcus pneumoniae	protein sequence ID NO:375.
KW	Streptococcus pneumoniae	strain 0100993; vaccine; immune response;
KW	Streptococcal infection;	pneumococcal.
OS	Streptococcus pneumoniae.	
PN	W09737026-A1.	
PD	09-OCT-1997.	

**THIS PAGE BLANK (USPTO)**

QY 2 PXXXXXXAVILM 14  
| :|:|

RESULT 13 PRELIMINARY; PRT; 815 AA.  
ID O81833;  
AC O81833;  
DT 01-NOV-1998 (TREMBlrel. 08, Created)  
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)  
DT 01-MAY-1999 (TREMBlrel. 10, Last annotation update)  
DE PUTATIVE RECEPTOR PROTEIN KINASE.  
GN M4122.110.  
OS Arabidopsis thaliana (Mouse-ear cross).  
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;  
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;  
OC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;  
OC Arabidopsis.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA REICHERT B.J., BAREL E., HOEISEL J., MEWES H.W., MAYER K.,  
RA SCHUELLER C., BEVAN M.;  
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA EU ARABIDOPSIS SEQUENCING PROJECT;  
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AL030978; CAAL9724.1; -;  
DR PFAM; PF00069; pkinase.1;  
DR PFAM; PF00954; S\_luciferase; 1;  
SQ SEQUENCE 815 AA; 91874 MW; CFB6E8A8 CRC32;

Query Match 91.1%; Score 41; DB 10; Length 815;  
Best Local Similarity 30.8%; Pred. No. 2.91e+01;  
Matches 4; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

Db 759 PEDRPTMASVILM 771  
| :|:|  
QY 2 PXXXXXXAVILM 14

RESULT 14 PRELIMINARY; PRT; 979 AA.  
ID O23377;  
AC O23377;  
DT 01-JAN-1998 (TREMBlrel. 05, Created)  
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)  
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)  
DE ABC TRANSPORTER-LIKE PROTEIN.  
GN DL3660W.  
OS Arabidopsis thaliana (Mouse-ear cross).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;  
OC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;  
OC Arabidopsis.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA BEVAN M., STIEREMA W., MURPHY G., WAMBUIT R., POHL T., TERRYN N.,  
RA KREIS M., KAVANAGH T., ENTIAN K.D., RIEGER M., JAMES R.,  
RA PUIGDOMENECH P., HATZOPOULOS P., OBERMAIER B., DUESTERHOFT A.,  
RA JONES J., PALME K., ANSGORE W., DELSENY M., BANCROFT I., MEWES H.W.,  
RA SCHUELLER C., CHALWATZIS N.;  
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA EU ARABIDOPSIS SEQUENCING PROJECT;  
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; Z97338; CAB45997.1; -;  
DR PFAM; PF00005; ABC\_tran; 1;  
SQ SEQUENCE 979 AA; 111606 MW; C8D713C9 CRC32;

Query Match 91.1%; Score 41; DB 10; Length 979;  
Best Local Similarity 38.5%; Pred. No. 2.91e+01;  
Matches 5; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Db 106 PETTELFDDVILM 118  
| :|:|  
QY 2 PXXXXXXAVILM 14

RESULT 15 PRELIMINARY; PRT; 1420 AA.  
ID O81016;  
AC O81016;  
DT 01-NOV-1998 (TREMBlrel. 08, Created)  
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)  
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)  
DE PUTATIVE ABC TRANSPORTER PROTEIN.  
GN F12C20.5.  
OS Arabidopsis thaliana (Mouse-ear cross).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;  
OC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;  
OC Arabidopsis.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA STRAIN=CV. COLUMBIA;  
RA ROUNDLEY S.D., RONNING C.M., LIN X., KETCHUM K.A., CROSBY M.L.,  
RA BRANDON R.C., SYKES S.M., KAUL S., MASON T.M., KERLAVAGE A.R.,  
RA ADAMS M.D., SOMERVILLE C.R., VENTER J.C.;  
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AC005168; AAC32236.1; -;  
DR MENDEL; 31671; Arabidopsis; 31671.  
DR PFAM; PF00005; ABC\_tran; 2;  
SQ SEQUENCE 1420 AA; 161264 MW; B8BC4333 CRC32;

Query Match 91.1%; Score 41; DB 10; Length 1420;  
Best Local Similarity 38.5%; Pred. No. 2.91e+01;  
Matches 5; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Db 369 PETTELFDDVILM 381  
| :|:|  
QY 2 PXXXXXXAVILM 14

Search completed: Sat Apr 15 01:30:11 2000  
Job time : 92 secs.

```
QY 2 PXXXXXXAVILM 14
RESULT 9
ID Q9VP16 PRELIMINARY; PRT; 383 AA.
AC Q9VP16
DT 01-MAY-1999 (TRENBLrel. 10, Created)
DT 01-MAY-1999 (TRENBLrel. 10, Last sequence update)
DT 01-MAY-1999 (TRENBLrel. 10, Last annotation update)
DE HYPOTHETICAL 43.8 KD PROTEIN.
OS Choristoneura fumiferana nuclear polyhedrosis virus (CfMNPV).
OC Viruses; dsDNA viruses, no RNA stage; Baculoviridae;
OC Nucleopolyhedrovirus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IRELAND;
RA POLOUMIENKO A., KRELL P.J.;
RT "Identification of the ORF 8/6 gene in the EGT-IAP1 intergenic region
of a baculovirus pathogenic to the spruce budworm, Cf MNPV.";
RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; U70432; AAD10307.1; -
KW Hypothetical protein.
SQ SEQUENCE 383 AA; 43752 MW; 81785249 CRC32;

Query Match 91.1%; Score 41; DB 14; Length 383;
Best Local Similarity 30.8%; Pred. No. 2.91e+01;
Matches 4; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

Db 136 PDCNRETSLVLM 148
QY 2 PXXXXXXAVILM 14
RESULT 10
ID Q9X3X9 PRELIMINARY; PRT; 487 AA.
AC Q9X3X9
DT 01-NOV-1999 (TRENBLrel. 12, Created)
DT 01-NOV-1999 (TRENBLrel. 12, Last sequence update)
DT 01-NOV-1999 (TRENBLrel. 12, Last annotation update)
DE TETRACENOMYCIN C RESISTANCE AND EXPORT PROTEIN.
GN YJCC.
OS Zymomonas mobilis.
OC Bacteria; Proteobacteria; alpha subdivision; Zymomonas group;
OC Zymomonas.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ZM4;
RA LEE H.J., KANG H.S.;
RT "Sequence analysis of 42C11 fosmid clone of Zymomonas mobilis ZM4.";
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF088896; AAD21553.1; -
SQ SEQUENCE 487 AA; 51863 MW; FFA132A6 CRC32;

Query Match 91.1%; Score 41; DB 2; Length 487;
Best Local Similarity 38.5%; Pred. No. 2.91e+01;
Matches 5; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

Db 364 PPHTKPIIAVINM 376
QY 2 PXXXXXXAVILM 14
RESULT 11
ID Q9ZP16 PRELIMINARY; PRT; 667 AA.
AC Q9ZP16
DT 01-MAY-1999 (TRENBLrel. 10, Created)
DT 01-MAY-1999 (TRENBLrel. 10, Last sequence update)
DT 01-NOV-1999 (TRENBLrel. 12, Last annotation update)
DE RECEPTOR-LIKE PROTEIN KINASE, RLK3 PRECURSOR.
GN RLK3.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;

core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
Arabidopsids.
[1]
SEQUENCE FROM N.A.
CZERNIC P., VISSER B., SUN W., SAVOURE A., DESLANDES L., MARCO Y.,
VAN MONTAGU M., VERBRUGGEN N.;
"Characterisation of an Arabidopsis thaliana receptor like protein
kinase gene activated by oxidative stress and pathogen attack.";
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ011674; CAA09731.1; -
DR MENDEL; 40122; Arath:1197;40122.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
KW Signal; Receptor; Kinase.
FT SIGNAL 1 25 POTENTIAL.
FT CHAIN 26 667 RECEPTOR-LIKE PROTEIN KINASE, RLK3.
SQ SEQUENCE 667 AA; 74081 MW; 595BD115 CRC32;

Query Match 91.1%; Score 41; DB 10; Length 667;
Best Local Similarity 30.8%; Pred. No. 2.91e+01;
Matches 4; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

Db 611 PEDRPLMLSTIILM 623
QY 2 PXXXXXXAVILM 14
RESULT 12
ID Q19444 PRELIMINARY; PRT; 684 AA.
AC Q19444
DT 01-NOV-1996 (TRENBLrel. 01, Created)
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
DT 01-NOV-1998 (TRENBLrel. 08, Last annotation update)
DE SIMILAR TO NA(+/H(+)) ANTIPOINTER.
GN F14B8.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;
OC Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE; 94150718.
RA WILSON R., AINSOUGH R., ANDERSON K., BAYNES C., BERKS M., COULSON A.,
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., FULTON L.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., JEROME L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN T., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
elegans.";
RL Nature 368:32-38(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA GEISEL C.;
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA WATERSTON R.;
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; U28737; AAA68274.1; -
DR PFAM; PF00999; Na_H_Exchange; 1.
SQ SEQUENCE 684 AA; 76375 MW; CE96AA60 CRC32;

Query Match 91.1%; Score 41; DB 5; Length 684;
Best Local Similarity 30.8%; Pred. No. 2.91e+01;
Matches 4; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

Db 56 PVLKRLPSVILM 68
```

DR PFAM; PF00528; BPD\_transp: 1.  
KW Hypothetical protein; Transport; Transmembrane.  
SQ SEQUENCE 223 AA; 24751 MW; 2D540CAB CRC32;

Query Match 93.3%; Score 42; DB 2; Length 223;  
Best Local Similarity 38.5%; Pred. No. 1.69e+01;  
Matches 5; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

Db 186 PNASFVYGVILM 198

QY 2 PXXXXXXAVILM 14

RESULT 6  
ID Q92038 PRELIMINARY; PRT; 292 AA.  
AC Q92038;  
DT 01-JAN-1999 (TREMELrel. 09, Created)  
DT 01-JAN-1999 (TREMELrel. 09, Last sequence update)  
DT 01-NOV-1999 (TREMELrel. 12, Last annotation update)  
DE ACYL-COA DESATURASE (EC 1.14.99.5) (STEAROYL-COA DESATURASE) (FATTY ACID DESATURASE) (DELTA(9)-DESATURASE).  
OS Cyprinus carpio (Common carp).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi; Cypriniformes; Cyprinidae; Cyprinidae; Cyprinidae; Cyprinus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=LIVER;  
RX MEDLINE; 96223338.  
RA TIKU P.E., GRACEY A.Y., MACARTNEY A.I., BRYNOR R.J., COSSINS A.R.; "Cold-induced expression of delta 9-desaturase in carp by transcriptional and posttranslational mechanisms."; Science 271:815-818 (1996).  
RL Science 271:815-818 (1996).  
CC -1- FUNCTION: THIS DELTA-9 DESATURASE IS A TERMINAL COMPONENT OF THE LIVER MICROSOMAL STEAROYL-COA DESATURASE SYSTEM, THAT UTILIZES O(2) AND ELECTRONS FROM REDUCED CYTOCHROME B(5) TO CATALYZE THE INSERTION OF A DOUBLE BOND INTO A SPECTRUM OF FATTY ACYL-COA SUBSTRATES (BY SIMILARITY).  
CC -1- CATALYTIC ACTIVITY: STEAROYL-COA + AH(2) + O(2) = OLEOYL-COA + A + 2 H(2)O.  
CC -1- COFACTOR: IRON (BY SIMILARITY).  
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. ENDOPLASMIC RETICULUM (PROBABLE).  
CC -1- INDUCTION: BY COLD. A 10-FOLD INCREASE IN TRANSCRIPT LEVELS IS OBSERVED 48-60 HOURS AFTER COOLING.  
CC -1- DOMAIN: THE HISTIDINE BOX DOMAINS MAY CONTAIN THE ACTIVE SITE AND/OR BE INVOLVED IN METAL ION BINDING.  
CC -1- SIMILARITY: TO OTHER FATTY ACID DESATURASES.  
DR EMBL; U31864; AAB03857.1; -.  
DR PFAM; PF01069; Desaturase\_1.  
DR PRINTS; PR00075; FACDSATASE.  
KW Oxidoreductase; Fatty acid biosynthesis; Iron; Transmembrane;  
KW Endoplasmic reticulum; Glycoprotein.  
FT TRANSMEM 47 67 POTENTIAL.  
FT TRANSMEM 69 89 POTENTIAL.  
FT TRANSMEM 203 223 POTENTIAL.  
FT DOMAIN 91 96 HISTIDINE BOX 1.  
FT DOMAIN 128 132 HISTIDINE BOX 2.  
FT DOMAIN 269 273 HISTIDINE BOX 3.  
FT CARBOHYD 27 27 POTENTIAL.  
FT CARBOHYD 230 230 POTENTIAL.  
FT CARBOHYD 289 289 POTENTIAL.  
SQ SEQUENCE 292 AA; 33649 MW; 72418D20 CRC32;

Query Match 91.1%; Score 41; DB 13; Length 292;  
Best Local Similarity 38.5%; Pred. No. 2.91e+01;  
Matches 5; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Db 38 PPTVIVWRNVILM 50

QY 2 PXXXXXXAVILM 14

RESULT 7  
ID Q28879 PRELIMINARY; PRT; 312 AA.  
AC Q28879;  
DT 01-JAN-1998 (TREMELrel. 05, Created)  
DT 01-JAN-1998 (TREMELrel. 05, Last sequence update)  
DT 01-AUG-1998 (TREMELrel. 07, Last annotation update)  
DE BRANCHED-CHAIN AMINO ACID ABC TRANSPORTER, PERMEASE PROTEIN (BRAD-4).  
GN AFI392.  
OS Archaeoglobus fulgidus.  
OC Archaea; Euryarchaeota; Archaeoglobales; Archaeoglobaceae;  
OC Archaeoglobus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;  
RX MEDLINE; 98049343.  
RA KLENK H.-P., CLAYTON R.A., TOMB J.-F., WHITE O., NELSON K.E., KETCHUM K.A., DOOSON R.J., GWINN M., HICKEY E.K., PETERSON J.D., RICHARDSON D.L., KERLAVAGE A.R., GRAHAM D.E., KYRPIDES N.C., FLEISCHMANN R.D., QUACKENBUSH J., LEE N.H., SUTTON G.G., GILL S., KIRKNESS E.F., DOUGHERTY B.A., MCKENNEY K., ADAMS M.D., LOFTUS B., PETERSON S., REICH C.I., MCNEIL L.K., BADGER J.H., GLODEK A., ZHOU L., OVERBECK R., GOCAYNE J.D., WEIDMAN J.F., McDONALD L., UTTERBACK T., COTTON M.D., SPRIGGS T., ARTIACH P., RAINE B.P., SYKES S.M., SADOW P.W., D'ANDREA K.P., BOWMAN C., FUJII C., GARLAND S.A., MASON T.M., OLSEN G.J., FRASER C.M., SMITH H.O., WOESE C.R., VENTER J.C.; "The complete genome sequence of the hyperthermophilic, sulphate-reducing archaeon Archaeoglobus fulgidus."; Nature 390:364-370 (1997).  
DR EMBL; AF001008; AAB89857.1; -.  
DR TIGR; AFI392; -.  
KW Hypothetical protein.  
SQ SEQUENCE 312 AA; 33392 MW; 8477F90E CRC32;

Query Match 91.1%; Score 41; DB 1; Length 312;  
Best Local Similarity 38.5%; Pred. No. 2.91e+01;  
Matches 5; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Db 92 PLIRGASVVILM 104

QY 2 PXXXXXXAVILM 14

RESULT 8  
ID Q95499 PRELIMINARY; PRT; 314 AA.  
AC Q95499;  
DT 01-MAY-1999 (TREMELrel. 10, Created)  
DT 01-MAY-1999 (TREMELrel. 10, Last sequence update)  
DT 01-MAY-1999 (TREMELrel. 10, Last annotation update)  
DE OLFACTORY RECEPTOR 89.  
GN OLF899.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA AYER LELIEVRE C., AMADOU C., GALLINARO H., AVOUSTIN P., RIBOUCHON M., BOUISO C., TAZI AHNINI R., PONTAROTI P.; "Olfactory receptor gene cluster in man and mouse major histocompatibility complex (MHC): New insights into the evolution of vertebrate olfactory receptor gene family."; Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AJ132194; CAA10602.1; -.  
KW Receptor.  
SQ SEQUENCE 314 AA; 35128 MW; 4669D1F8 CRC32;

Query Match 91.1%; Score 41; DB 4; Length 314;  
Best Local Similarity 30.8%; Pred. No. 2.91e+01;  
Matches 4; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

Db 36 PIAVMGNITILM 48

QY 1



ID P77007 PRELIMINARY; PRT; 373 AA.  
AC F77007;  
DT 01-FEB-1997 (TREMELrel. 02, Created)  
DT 01-FEB-1997 (TREMELrel. 02, Last sequence update)  
DT 01-NOV-1999 (TREMELrel. 12, Last annotation update)  
DE SIGNAL RECOGNITION PARTICLE PROTEIN (FIFTY-FOUR HOMOLOG) (P48)  
DE (FRAGMENT).  
GN FFH.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Escherichia.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-K12;  
RX MEDLINE; 97349980.  
RA YAMAMOTO Y., AIBA H., BABA T., HAYASHI K., INADA T., ISONO K.,  
RA YAMAMOTO Y., KIMURA S., KITAGAWA M., MAKINO K., MIKI T., MITSUHASHI N.,  
RA MIZOBUCHI K., MORI H., NAKADE S., NAKAMURA Y., NASHIMOTO H.,  
RA OSHIMA T., OYAMA S., SATO N., SAMPEI G., SATOH Y., SIVASUNDARAM S.,  
RA TAGAMI H., TAKAHASHI H., TAKEDA J., TAKEMOTO K., UEHARA K., WADA C.,  
RA YAMAGATA S., HORIUCHI T.;  
RT "Construction of a contiguous 874-kb sequence of the Escherichia coli  
RT -K12 genome corresponding to 50.0-68.8 min on the linkage map and  
RT analysis of its sequence features";  
RL DNA Res. 4:91-113(1997).  
DR EMBL; D90888; BA016495.1; -.  
DR HSSP; O07347; 1FFH.  
DR PROSITE; P500300; SRP54; 1.  
DR PFAM; PF00448; SRP54; 1.  
FT NON\_TER 1  
SQ SEQUENCE 373 AA; 40577 MW; 845E5D8E CRC32;

Query Match 95.6%; Score 43; DB 2; Length 373;  
Best Local Similarity 38.5%; Pred. No. 9.59e+00;  
Matches 5; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

Db 13 PEFGCATAAVLVM 25  
QY 2 PXXXXXXAVILM 14

RESULT 3  
ID O92389 PRELIMINARY; PRT; 382 AA.  
AC O92389;  
DT 01-NOV-1998 (TREMELrel. 08, Created)  
DT 01-NOV-1998 (TREMELrel. 08, Last sequence update)  
DT 01-NOV-1998 (TREMELrel. 08, Last annotation update)  
DE ACMNPV ORF22.  
OS Bombyx mori nuclear polyhedrosis virus (BmNPV).  
OC Viruses; dsDNA viruses, no RNA stage; Baculoviridae;  
OC Nucleopolyhedrovirus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-T3;  
RX MEDLINE; 97329351.  
RA KAMITA S.G., MAEDA S.;  
RT "Sequencing of the putative DNA helicase-encoding gene of the Bombyx  
RT mori nuclear polyhedrosis virus and fine-mapping of a region involved  
RT in host range expansion";  
RL Gene 190:173-179(1997).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-T3;  
RX GOMI S., MAJIMA K., MAEDA S.;  
RT "Sequence analysis of the genome of Bombyx mori  
RT nucleopolyhedrovirus";  
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL; L33180; AAC63696.1; -.  
SQ SEQUENCE 382 AA; 43815 MW; 06CC3A6A CRC32;

Query Match 95.6%; Score 43; DB 14; Length 382;  
Best Local Similarity 38.5%; Pred. No. 9.59e+00;  
Matches 5; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

Db 136 PRNCNRETSVILM 148  
QY 2 PXXXXXXAVILM 14

RESULT 4  
ID O92K45 PRELIMINARY; PRT; 223 AA.  
AC O92K45;  
DT 01-MAY-1999 (TREMELrel. 10, Created)  
DT 01-MAY-1999 (TREMELrel. 10, Last sequence update)  
DT 01-NOV-1999 (TREMELrel. 12, Last annotation update)  
DE AMINO ACID ABC TRANSPORTER, PERMEASE PROTEIN.  
GN JHP1097.  
OS Helicobacter pylori J99.  
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;  
OC Helicobacter.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-J99;  
RX MEDLINE; 99120557.  
RA ALM R.A., LING L.-S.L., MOIR D.T., KING B.L., BROWN E.D., DOIG P.C.,  
RA SMITH D.R., NOONAN B., GUILD B.C., DEJONGE B.L., CARMEL G.,  
RA TUMMINO P.J., CARUSO A., URIA-NICKELSEN M., MILLS D.M., IVES C.,  
RA GIBSON R., MERBERG D., MILLS S.D., JIANG Q., TAYLOR D.E., VOVIS G.F.,  
RA TRUST T.J.;  
RT "Genomic-sequence comparison of two unrelated isolates of the human  
RT gastric pathogen Helicobacter pylori";  
RL Nature 397:176-180(1999).  
DR EMBL; AE001537; AAD06679.1; -.  
DR PROSITE; P500402; BPD\_TRANSF\_INN\_MEMBER; 1.  
KW Transport; Transmembrane.  
SQ SEQUENCE 223 AA; 24749 MW; ABD453CD CRC32;

Query Match 93.3%; Score 42; DB 2; Length 223;  
Best Local Similarity 38.5%; Pred. No. 1.68e+01;  
Matches 5; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

Db 186 PNASFVYGVILM 198  
QY 2 PXXXXXXAVILM 14

RESULT 5  
ID O25784 PRELIMINARY; PRT; 223 AA.  
AC O25784;  
DT 01-JAN-1998 (TREMELrel. 05, Created)  
DT 01-JAN-1998 (TREMELrel. 05, Last sequence update)  
DT 01-NOV-1999 (TREMELrel. 12, Last annotation update)  
DE GLUTAMINE ABC TRANSPORTER, PERMEASE PROTEIN (GLNP).  
GN HPI170.  
OS Helicobacter pylori (Campylobacter pylori).  
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;  
OC Helicobacter.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-26695;  
RX MEDLINE; 97394467.  
RA TOMB J.-F., WHITE O., KERLAVAGE A.R., CLAYTON R.A., SUTTON G.G.,  
RA FLEISCHMANN R.D., KETCHUM K.A., KLENK H.-P., GILL S., DOUGHERTY B.A.,  
RA NELSON K., QUACKENBUSH J., ZHOU L., KIRKNESS E.F., PETERSON S.,  
RA LOFTUS B., RICHARDSON D., DODSON R., KHALAK H.G., GLODEK A.,  
RA MCKENNEY K., FITZGERALD L.M., LEE N., ADAMS M.D., HICKEY E.K.,  
RA BERG D.E., GOCAYNE J.D., UTTERBACK T.R., PETERSON J.D., KELLEY J.M.,  
RA COTTON M.D., WEIDMAN J.M., FUJII C., BOWMAN C., WATTHEY L., WALLIN E.,  
RA HAYES W.S., BORODOVSKY M., KARP P.D., SMITH H.O., FRASER C.M.,  
RA VENTER J.C.;  
RT "The complete genome sequence of the gastric pathogen Helicobacter  
RT pylori";  
RL Nature 388:539-547(1997).  
DR EMBL; AE000623; AAD08216.1; -.  
DR TIGR; HPI170; -.  
DR PROSITE; P500402; BPD\_TRANSF\_INN\_MEMBER; 1.

\*\*\*\*\*

WATER

(TM)

\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Sat Apr 15 01:28:39 2000; MasPar time 7.60 Seconds  
127.722 Million cell updates/sec

Tabular output not generated.

Title: >US-08-452-843-23  
Description: (1-14) from US08452843.pep  
Perfect Score: 45  
Sequence: 1 XPXXXXXXAVILM 14

Scoring table: PAM 150  
Gap 15

Searched: 225878 seqs, 69334122 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: sptrmbl12

1:sp\_archaea 2:sp\_bacteria 3:sp\_fungi 4:sp\_human  
5:sp\_invertebrate 6:sp\_mammal 7:sp\_mhc 8:sp\_organelle  
9:sp\_phase 10:sp\_plant 11:sp\_rodent 12:sp\_unclassified  
13:sp\_vertebrate 14:sp\_virus

Statistics: Mean 20.232; Variance 20.398; scale 0.992

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	45	100.0	485	10	AUX1 GENE.	3.04e+00
2	43	95.6	373	2	SIGNAL RECOGNITION PAR	9.59e+00
3	43	95.6	382	14	ACMPV ORF22.	9.59e+00
4	42	93.3	223	2	AMINO ACID ABC TRANSP	1.68e+01
5	42	93.3	223	2	GLUTAMINE ABC TRANSP	1.68e+01
6	41	91.1	292	13	ACYL-COA DESATURASE (E	2.91e+01
7	41	91.1	312	1	BRANCHED-CHAIN AMINO A	2.91e+01
8	41	91.1	314	4	OLFACTORY RECEPTOR 89	2.91e+01
9	41	91.1	383	14	HYPOTHETICAL 43.8 KD P	2.91e+01
10	41	91.1	487	2	TEPEENOMYCIN C RESIS	2.91e+01
11	41	91.1	667	10	RECEPTOR-LIKE PROTEIN	2.91e+01
12	41	91.1	684	5	SIMILAR TO NA(+/H(+)	2.91e+01
13	41	91.1	815	10	PUTATIVE RECEPTOR PROT	2.91e+01
14	41	91.1	979	10	ABC TRANSPORTER-LIKE P	2.91e+01
15	41	91.1	1420	10	PUTATIVE ABC TRANSPORT	2.91e+01
16	40	88.9	280	2	SUGAR TRANSPORT PROTEI	5.00e+01
17	40	88.9	340	10	F11O4.14 PROTEIN.	5.00e+01
18	40	88.9	351	2	YFKE PROTEIN.	5.00e+01
19	40	88.9	400	4	GDNF FAMILY RECEPTOR A	5.00e+01
20	40	88.9	408	2	DELTA-1-PYRROLINE-5-CA	5.00e+01

21	40	88.9	411	5	Q20414	F4G4.4 PROTEIN.	5.00e+01
22	40	88.9	416	3	Q9Y797	INOSITOLPHOSPHORYLCERA	5.00e+01
23	40	88.9	444	10	Q9Z0E2	PUTATIVE REVERSE TRANS	5.00e+01
24	40	88.9	452	5	Q18924	SIMILAR TO ARYL SULFATA	5.00e+01
25	40	88.9	489	1	Q9YA20	489AA LONG HYPOTHETICA	5.00e+01
26	40	88.9	637	5	Q44547	R02C2.1 PROTEIN.	5.00e+01
27	40	88.9	971	13	Q42573	NEDD4 PROTEIN.	5.00e+01
28	40	88.9	995	4	Q9X165	KCSABII-B PROTEIN.	5.00e+01
29	40	88.9	1518	2	Q9WX75	KIAA0439 (FRAGMENT).	5.00e+01
30	39	86.7	112	2	P76307	FROM BASES 1975196 TO	8.49e+01
31	39	86.7	252	7	Q78197	HISTOCOMPATIBILITY 2.	8.49e+01
32	39	86.7	263	7	Q31184	H2-IA-BETA CELL SURFAC	8.49e+01
33	39	86.7	263	7	Q31135	HISTOCOMPATIBILITY 2.	8.49e+01
34	39	86.7	263	7	Q31187	MHC CLASS II H2-IA-BET	8.49e+01
35	39	86.7	265	7	Q31131	MHC CLASS II H-2 I-A B	8.49e+01
36	39	86.7	359	4	Q9Y695	ACYL-COA DESATURASE 1	8.49e+01
37	39	86.7	359	6	Q62849	ACYL-COA DESATURASE 1	8.49e+01
38	39	86.7	379	10	Q9X255	HYPOTHETICAL 40.2 KD P	8.49e+01
39	39	86.7	446	2	Q9XAF1	PUTATIVE INTEGRAL MEMB	8.49e+01
40	39	86.7	472	5	Q9X255	N-MYRISTOYL TRANSFERAS	8.49e+01
41	39	86.7	633	10	Q65470	SERINE/THREONINE KINAS	8.49e+01
42	39	86.7	861	2	Q9X516	PUTATIVE ACONITATE HYD	8.49e+01
43	39	86.7	944	10	Q48538	RSOHAP108.	8.49e+01
44	39	86.7	944	10	Q80342	ATRBOH F PROTEIN.	8.49e+01
45	39	86.7	1124	10	Q49318	PUTATIVE RECEPTOR PROT	8.49e+01

ALIGNMENTS

RESULT	ID	PRELIMINARY;	PRT;	485 AA.
1	Q96247	DT 01-FEB-1997 (TREMELrel. 02, Created)		
2	Q96247	DT 01-FEB-1997 (TREMELrel. 02, Last sequence update)		
3	Q96247	DT 01-NOV-1999 (TREMELrel. 12, Last annotation update)		
4	Q96247	DE AUX1 GENE.		
5	Q96247	OS Arabidopsis thaliana (Mouse-ear cross).		
6	Q96247	GN Arabidopsis thaliana (Mouse-ear cross).		
7	Q96247	OC Eukaryota: Viridiplantae; Streptophyta; Tracheophyta;		
8	Q96247	OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;		
9	Q96247	OC core eudicots; Rosidae; euroids II; Brassicales; Brassicaceae;		
10	Q96247	OC Arabidopsis.		
11	Q96247	RN [1]		
12	Q96247	RP SEQUENCE FROM N.A.		
13	Q96247	RA MEDLINE; 96337989.		
14	Q96247	RA BENNETT M.J., MARCHANT A., GREEN H.G., MAY S.T., WARD S.P.,		
15	Q96247	RA MILLNER P.A., WALKER A.R., SCHULTZ B., FELDMANN K.A.;		
16	Q96247	RT "Arabidopsis AUX1 gene: a permease-like regulator of root		
17	Q96247	RT gravitropism."		
18	Q96247	RL Science 273:948-950(1996).		
19	Q96247	RN [2]		
20	Q96247	RP SEQUENCE FROM N.A.		
21	Q96247	RC STRAIN=CV. COLUMBIA;		
22	Q96247	RA ROUNSLEY S.D., KAUL S., LIN X., KETCHUM K.A., CROSBY M.L.,		
23	Q96247	RA BRANDON R.C., SYKES S.M., MASON T.M., KERLAVAGE A.R., ADAMS M.D.,		
24	Q96247	RA SOMERVILLE C.R., VENTER J.C.;		
25	Q96247	RT "Arabidopsis thaliana chromosome II BAC F16M14 genomic sequence."		
26	Q96247	RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.		
27	Q96247	DR EMBL; X98772; CAA67308.1; -		
28	Q96247	DR EMBL; AC003028; AAC27161.1; -		
29	Q96247	DR PFAM; PF01490; Aa_trans; 1.		
30	Q96247	SQ SEQUENCE 485 AA; 54059 MW; 683DA584 CRC32;		

Query Match	100.08;	Score 45;	DB 10;	Length 485;
Best Local Similarity	46.2%;	Pred. No. 3.04e+00;		
Matches	6;	Conservative 0;	Mismatches 7;	Indels 0; Gaps 0;
Db	306	PKNAWRDAAVILM 318		
Qy	2	PXXXXXXAVILM 14		
RESULT	2			

**THIS PAGE BLANK (USPTO)**

```
RESULT 15
ID HB2Q_MOUSE STANDARD; PRT; 265 AA.
AC P06342;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE H-2 CLASS II HISTOCOMPATIBILITY ANTIGEN, A-Q BETA CHAIN PRECURSOR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 86233280.
RA ESTESS P., BEGOVICH A.B., KOO M., JONES P.P., MCDEVITT H.O.;
RT "Sequence analysis and structure-function correlations of murine q,
RT k, u, s, and f haplotype I-A beta cDNA clones.";
RL Proc. Natl. Acad. Sci. U.S.A. 83:3594-3598(1986).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M13537; AAA39633.1; -.
DR PIR; A02237; HLMSQB.
DR HSP; P06343; IIAK.
DR PROSITE; PS00290; IG_MHC; 1.
DR PFAM; PF00047; Ig; 1.
DR PFAM; PF00969; MHC_II_beta; 1.
KW MHC II; Transmembrane; Glycoprotein; Signal.
FT SIGNAL 1 27
FT CHAIN 28 265 H-2 CLASS II HISTOCOMPATIBILITY ANTIGEN,
FT A-Q BETA CHAIN.
FT DOMAIN 28 122 EXTRACELLULAR BETA-1.
FT DOMAIN 123 217 EXTRACELLULAR BETA-2.
FT DOMAIN 218 227 CONNECTING PEPTIDE.
FT TRANSMEM 228 247
FT DOMAIN 248 265 CYTOPLASMIC TAIL.
FT DISULFID 42 106 BY SIMILARITY.
FT DISULFID 145 201 BY SIMILARITY.
FT CARBOHYD 46 46 POTENTIAL.
SQ SEQUENCE 265 AA; 29950 MW; C02E023F CRC32;
```

```
Query Match 86.7%; Score 39; DB 1; Length 265;
Best Local Similarity 30.8%; Pred. No. 3.30e+01;
Matches 4; Conservative 1; Mismatches 8; Indels 0; Gaps 0;
```

```
Db 6 PSLLLSAAVVVLM 18
QY 2 PXXXXXXAVILM 14
```

```
Search completed: Sat Apr 15 01:28:22 2000
Job time : 42 secs.
```

```

FT DOMAIN 248 265 CYTOPLASMIC TAIL.
FT DISULFID 42 106 BY SIMILARITY.
FT DISULFID 143 201 BY SIMILARITY.
FT CARBOHYD 46 46 POTENTIAL.
SQ: SEQUENCE 265 AA; 30128 MW; E4C6D72D CRC32;

Query Match 86.7%; Score 39; DB 1; Length 265;
Best Local Similarity 30.8%; Pred. No. 3.30e+01;
Matches 4; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

Db 6 PSLLLSAAVVVLM 18
| | | |
QY 2 PXXXXXXXAVILM 14

RESULT 14
ID HB2D_MOUSE STANDARD: PRT; 265 AA.
AC F01921; O19458; Q31138; Q31139;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-DEC-1999 (Rel. 39, Last annotation update)
DE H-2 CLASS II HISTOCOMPATIBILITY ANTIGEN, A-D BETA CHAIN PRECURSOR.
GN H2-AB1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BALB/C;
RX MEDLINE; 83275703.
RA MALLISSEN M., HUNKAPILLER T., HOOD L.E.;
RT "Nucleotide sequence of a light chain gene of the mouse I-A
subregion: A beta d.";
RL Science 221:750-754(1983).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; K00008; AAA39548.1; .
CC EMBL; K00007; AAA39548.1; JOINED.
CC PIR; A02236; HLMSAB.
CC HSSP; P06343; IIAK.
CC MGD; MGI:103070; H2-Ab1.
CC PROSITE; PS00290; IG_MHC; 1.
CC PFAM; PF00047; ig; 1.
CC PFAM; PF00969; MHC_II_beta; 1.
CC MHC II; Transmembrane; Glycoprotein; Signal.
KW SIGNAL 1 27
FT CHAIN 28 265 H-2 CLASS II HISTOCOMPATIBILITY ANTIGEN,
FT FT A-D BETA CHAIN.
FT DOMAIN 28 122 EXTRACELLULAR BETA-1.
FT DOMAIN 123 216 EXTRACELLULAR BETA-2.
FT DOMAIN 217 226 CONNECTING PEPTIDE.
FT TRANSMEM 227 247
FT DOMAIN 248 265 CYTOPLASMIC TAIL.
FT DISULFID 42 106 BY SIMILARITY.
FT DISULFID 145 201 BY SIMILARITY.
FT CARBOHYD 46 46 POTENTIAL.
SQ SEQUENCE 265 AA; 29954 MW; 7085AC51 CRC32;

Query Match 86.7%; Score 39; DB 1; Length 265;
Best Local Similarity 30.8%; Pred. No. 3.30e+01;
Matches 4; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

Db 6 PSLLLSAAVVVLM 18
| | | |
QY 2 PXXXXXXXAVILM 14

```

```
DR PROSITE; PS01086; RIBUL_P3_EPIMER_2; 1.
KW isomerase; Carbohydrate metabolism.
SQ SEQUENCE 229 AA; 24987 MW; 2B7B2558 CRC32;

Query Match      86.78; Score 39; DB 1; Length 229;
Best Local Similarity 30.8; Pred. No. 3.30e+01;
Matches 4; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

Db 129 PSLPFCDDVVVLM 141
|
| 1:11
QY 2 PXXXXXXAVILM 14

RESULT 10
ID HB2S_MOUSE STANDARD; PRT; 263 AA.
AC P06345;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE H-2 CLASS II HISTOCOMPATIBILITY ANTIGEN, A-S BETA CHAIN PRECURSOR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 86233280.
RA ESTESS P., BEGOVICH A.B., KOO M., JONES P.P., MCDEVITT H.O.;
RT "Sequence analysis and structure-function correlations of murine q,
RT k, u, s, and f haplotype I-A beta cDNA clones.";
RL Proc. Natl. Acad. Sci. U.S.A. 83:3594-3598(1986).
DR PIR; A02240; HLMSBS.
DR HSSP; P06343; IG_MHC; 1.
DR PFAM; PF00047; Ig; 1.
DR PFAM; PF00969; MHC_II_beta; 1.
DR PFAM; PF00969; MHC_II_beta; 1.
KW MHC II; Transmembrane; Glycoprotein; Signal.
FT SIGNAL 1 27
FT CHAIN 28 263
FT DOMAIN 28 120 H-2 CLASS II HISTOCOMPATIBILITY ANTIGEN,
FT DOMAIN 121 214 A-S BETA CHAIN.
FT DOMAIN 215 224 EXTRACELLULAR BETA-1.
FT DOMAIN 225 245 EXTRACELLULAR BETA-2.
FT TRANSMEM 225 245 CONNECTING PEPTIDE.
FT DOMAIN 246 263 CYTOPLASMIC TAIL.
FT DISULFID 42 104 BY SIMILARITY.
FT DISULFID 143 199 BY SIMILARITY.
FT CARBOHYD 46 46 POTENTIAL.
SQ SEQUENCE 263 AA; 25788 MW; 7C8D17F1 CRC32;

Query Match      86.78; Score 39; DB 1; Length 263;
Best Local Similarity 30.8; Pred. No. 3.30e+01;
Matches 4; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

Db 6 PSLLSAAVVVLM 18
|
| 1:11
QY 2 PXXXXXXAVILM 14

RESULT 11
ID HB2U_MOUSE STANDARD; PRT; 263 AA.
AC P06344;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
```

```
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE H-2 CLASS II HISTOCOMPATIBILITY ANTIGEN, A-U BETA CHAIN PRECURSOR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 86233280.
RA ESTESS P., BEGOVICH A.B., KOO M., JONES P.P., MCDEVITT H.O.;
RT "Sequence analysis and structure-function correlations of murine q,
RT k, u, s, and f haplotype I-A beta cDNA clones.";
RL Proc. Natl. Acad. Sci. U.S.A. 83:3594-3598(1986).
DR PIR; A02239; HLMSBU.
DR HSSP; P06343; IIAK.
DR PFAM; PS00290; IG_MHC; 1.
DR PFAM; PF00047; Ig; 1.
DR PFAM; PF00969; MHC_II_beta; 1.
DR PFAM; PF00969; MHC_II_beta; 1.
KW MHC II; Transmembrane; Glycoprotein; Signal.
FT SIGNAL 1 27
FT CHAIN 28 263
FT DOMAIN 28 120 H-2 CLASS II HISTOCOMPATIBILITY ANTIGEN,
FT DOMAIN 121 214 A-U BETA CHAIN.
FT DOMAIN 215 224 EXTRACELLULAR BETA-1.
FT DOMAIN 225 245 EXTRACELLULAR BETA-2.
FT TRANSMEM 225 245 CONNECTING PEPTIDE.
FT DOMAIN 246 263 CYTOPLASMIC TAIL.
FT DISULFID 42 104 BY SIMILARITY.
FT DISULFID 143 199 BY SIMILARITY.
FT CARBOHYD 46 46 POTENTIAL.
SQ SEQUENCE 263 AA; 30041 MW; 4B80B069 CRC32;

Query Match      86.7%; Score 39; DB 1; Length 263;
Best Local Similarity 30.8; Pred. No. 3.30e+01;
Matches 4; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

Db 6 PSLLLAAVVVLM 18
|
| 1:11
QY 2 PXXXXXXAVILM 14

RESULT 12
ID HB2B_RAT STANDARD; PRT; 263 AA.
AC P29826;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE RT1 CLASS II HISTOCOMPATIBILITY ANTIGEN, B-1 BETA CHAIN PRECURSOR
DE (RT1.B-BETA(1)).
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-LEWIS; TISSUE-BONE MARROW;
RX MEDLINE; 91316148.
RA SYHA-JEDELHAUSER J., WENDLING U., RESKE K.;
RT "Complete coding nucleotide sequence of cDNA for the class II RT1.B
RT beta 1 chain of the Lewis rat.";
RL Biochim. Biophys. Acta 1089:414-416(1991).
CC -!- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO
CC THE IMMUNE SYSTEM.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X56596; CAA39934.1; -.
CC PIR; S18999; HLRTEB.
CC HSSP; P06343; IIAK.
```

Best Local Similarity 30.8%; Pred. No. 1.90e+01;  
Matches 4; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

Db 76 PPHYTTTPIILM 88  
Qy 2 PXXXXXXAVILM 14

RESULT 7  
ID VA53\_VACCV STANDARD; PRT; 103 AA.  
AC 24756; (Rel. 21, Created)  
DT 01-MAR-1992 (Rel. 21, Last sequence update)  
DT 01-MAR-1992 (Rel. 21, Last sequence update)  
DT 01-JUN-1994 (Rel. 29, Last annotation update)  
DE PROTEIN A53.  
GN A53R OR SALF16R OR SALF19R.  
OS Vaccinia virus (strain WR).  
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;  
CC Orthopoxvirus.  
[1]  
RN SEQUENCE FROM N.A.  
RP MEDLINE: 91259063.  
RA SMITH G.L., CHAN Y.S., HOWARD S.T.;  
RT "Nucleotide sequence of 42 kbp of vaccinia virus strain WR from near  
the right inverted terminal repeat.";  
RL J. Gen. Virol. 72:1349-1376(1991).  
[2]  
RN SEQUENCE FROM N.A.  
RP MEDLINE: 91111982.  
RA HOWARD S.T., CHAN Y.S., SMITH G.L.;  
RT "Vaccinia virus homologues of the Shope fibroma virus inverted  
terminal repeat proteins and a discontinuous ORF related to the tumor  
neurosis factor receptor family.";  
RL Virology 180:633-647(1991).  
CC -1- SIMILARITY: CONTAINS A LA-NGFR/TNFR-TYPE CYSTEINE-RICH REGION.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
CC EMBL; D11079; BAA01827.1; -  
CC EMBL; M58054; AAA48339.1; -  
DR PIR: B38550; B38550.  
DR PIR: J01791; J01791.  
DR PROSITE; PS00652; TNFR\_NGFR\_1; 1.  
DR PFAM; PF00020; TNFR\_C6; 1.  
DR LATE PROTEIN; Repeat.  
FT DOMAIN 36 103  
FT REPEAT 36 73  
FT REPEAT 74 103  
SQ SEQUENCE 103 AA; 12001 MW; F21E927D CRC32;  
Query Match 88.9%; Score 40; DB 1; Length 103;  
Best Local Similarity 30.8%; Pred. No. 1.90e+01;  
Matches 4; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

Db 76 PPHYTTTPIILM 88  
Qy 2 PXXXXXXAVILM 14

RESULT 8  
ID YIC9\_YEAST STANDARD; PRT; 142 AA.  
AC P40538;  
DT 01-FEB-1995 (Rel. 31, Created)  
DT 01-FEB-1995 (Rel. 31, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE HYPOTHETICAL 16.6 KD PROTEIN IN SSM4-IRK1 INTERGENIC REGION.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
CC EMBL; A001604; AAD18338.1; -  
CC PROSITE; PS01085; RIBUL\_P3-EPIMER\_1; 1.  
-----

GN YIL029C.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;  
CC Saccharomycetaceae; Saccharomycetes.  
[1]  
RN SEQUENCE FROM N.A.  
RP STRAIN=S288C / AB972;  
RA BARRELL B.G., BADCOCK K., BANKIER A.T., BOWMAN S., BROWN D.,  
RA CHURCHER C.M., CONNOR R., COPSEY T., DEAR S., DEVLIN K., FRASER A.,  
RA GENTLES S., HAMLYN N., HORSNELL T.S., HUNT S., JAGELS K., JONES M.,  
RA LOUIS E., LYE G., MOULE S., MOULE T., ODELL C., PEARSON D.,  
RA RAJANDREAM M.A., RILES L., ROWLEY N., SKELTON J., SMITH V.,  
RA WALSH S.V., WHITEHEAD S.;  
RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.  
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
CC EMBL; Z46881; CAA86962.1; -  
CC DR HYPOTHETICAL protein; Transmembrane.  
FT TRANSMEM 3 23 POTENTIAL.  
FT TRANSMEM 30 50 POTENTIAL.  
FT TRANSMEM 91 111 POTENTIAL.  
SQ SEQUENCE 142 AA; 16587 MW; 5559EEC2 CRC32;  
Query Match 86.7%; Score 39; DB 1; Length 142;  
Best Local Similarity 30.8%; Pred. No. 3.30e+01;  
Matches 4; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

Db 127 PGRVILSVILM 139  
Qy 2 PXXXXXXAVILM 14

RESULT 9  
ID REF\_CHLPN STANDARD; PRT; 229 AA.  
AC Q92829;  
DT 15-DEC-1999 (Rel. 39, Created)  
DT 15-DEC-1999 (Rel. 39, Last sequence update)  
DT 15-DEC-1999 (Rel. 39, Last annotation update)  
DE RIBULOSE-PHOSPHATE 3-EPIMERASE (EC 5.1.3.1) (PENTOSE-5-PHOSPHATE 3-  
DE EPIMERASE) (PPE) (R5P3E).  
GN RPE.  
OS Chlamydia pneumoniae (Chlamydophila pneumoniae).  
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydophila.  
[1]  
RN SEQUENCE FROM N.A.  
RP STRAIN=CWL029;  
RC MEDLINE: 99206606.  
RA KALMAN S., MITCHELL W., MARATHE R., LAMMEL C., FAN J., HYMAN R.W.,  
RA OLINGER L., GRIMWOOD J., DAVIS R.W., STEPHENS R.S.;  
RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";  
RL Nat. Genet. 21:385-389(1999).  
CC -1- CATALYTIC ACTIVITY: D-RIBULOSE 5-PHOSPHATE - D-XYLULOSE 5-  
CC PHOSPHATE.  
CC -1- SIMILARITY: BELONGS TO THE RIBULOSE-PHOSPHATE 3-EPIMERASE FAMILY.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
CC EMBL; AE001604; AAD18338.1; -  
CC PROSITE; PS01085; RIBUL\_P3-EPIMER\_1; 1.  
-----

```
RT "Isolation and characterization of a SCTL gene which can suppress a
RT choline-transport mutant of Saccharomyces cerevisiae."
RL J. Biochem. 117:447-451(1995).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN-S288C;
RX MEDLINE: 93070615.
RA SKALA J., VAN DYCK L., PURNELLE B., GOFFEAU A.;
RT "The sequence of an 8 kb segment on the left arm of chromosome II
RT from Saccharomyces cerevisiae identifies five new open reading frames
RT of unknown functions, two tRNA genes and two transposable elements."
RL Yeast 8:777-785(1992).
RN [3]
RP SEQUENCE OF 609-759 FROM N.A.
RX STRAIN-S288C;
RX MEDLINE: 93070613.
RA DELAVEAU T., JACO C., PEREA J.;
RT "Sequence of a 12.7 kb segment of yeast chromosome II identifies a
RT PDR-like gene and several new open reading frames."
RL Yeast 8:761-768(1992).
CC - FUNCTION: MULTICOPY SUPPRESSOR OF THE CTR1 MUTATION.
CC - SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (PROBABLE).
CC - SIMILARITY: STRONG, TO YEAST YKR067W.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: D38256; BAA07409.1; -
DR EMBL: 235773; CA84831.1; -
DR EMBL: 247695; CAB29349.1; -
DR PIR: S25330; S25330.
DR SGD: L0001820; SCTL.
KW Transmembrane.
FT TRANSMEM 49 66 POTENTIAL.
FT TRANSMEM 123 139 POTENTIAL.
FT TRANSMEM 260 276 POTENTIAL.
FT TRANSMEM 440 463 POTENTIAL.
FT TRANSMEM 494 516 POTENTIAL.
FT TRANSMEM 524 545 POTENTIAL.
FT DOMAIN 736 753 POLY-GLU.
FT CONFLICT 10 10 S -> F (IN REF. 2).
FT CONFLICT 88 88 R -> A (IN REF. 2).
FT CONFLICT 125 125 A -> P (IN REF. 2).
FT CONFLICT 730 730 S -> G (IN REF. 2 AND 3).
SQ SEQUENCE 759 AA; 85723 MW; BF15DFE5 CRC32;

Query Match 93.3%; Score 42; DB 1; Length 759;
Best Local Similarity 38.5%; Pred. No. 6.12e+00;
Matches 5; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

Db 89 PHANOFVDPVILM 101
QY 2 PXXXXXXAVILM 14

RESULT 5
ID Y022 NPVOP STANDARD; PRT; 382 AA.
AC 010281;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE HYPOTHETICAL 43.4 KD PROTEIN (ORF20).
OS Orgyia pseudotsugata multicapsid polyhedrosis virus (OpMNPV).
OC Viruses; dsDNA viruses, no RNA stage; Baculoviridae;
OC Nucleopolyhedrovirus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 97271300.
```

```
RA AHRENS C.H., RUSSELL R.R., FUNK C.J., EVANS J., HARWOOD S.,
RA ROHRMANN G.F.;
RT "The sequence of the Orgyia pseudotsugata multinucleocapsid nuclear
RT polyhedrosis virus genome."
RL Virology 229:381-393(1997).
CC - SIMILARITY: TO CORRESPONDING ORF IN ACNMPV.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U75930; AAC59019.1; -
DR EMBL: 382 AA; 43394 MW; 9EC81ECD CRC32;
KW Hypothetical protein.
SQ SEQUENCE 382 AA; 43394 MW; 9EC81ECD CRC32;

Query Match 91.1%; Score 41; DB 1; Length 382;
Best Local Similarity 30.8%; Pred. No. 1.09e+01;
Matches 4; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

Db 136 PRDCNRETSVULM 148
QY 2 PXXXXXXAVILM 14

RESULT 6
ID VA53 VACCC STANDARD; PRT; 103 AA.
AC P21071;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last annotation update)
DE PROTEIN A53.
GN A53R.
OS Vaccinia virus (strain Copenhagen).
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC Orthopoxvirus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 91021027.
RA GOEBEL S.J., JOHNSON G.P., PERKUS M.E., DAVIS S.W., WINSLOW J.P.,
RA PAOLETTI E.;
RT "The complete DNA sequence of vaccinia virus."
RL Virology 179:247-266(1990).
RN [2]
RP COMPLETE GENOME.
RA GOEBEL S.J., JOHNSON G.P., PERKUS M.E., DAVIS S.W., WINSLOW J.P.,
RA PAOLETTI E.;
RL Virology 179:517-563(1990).
CC - SIMILARITY: CONTAINS A LA-NGFR/TNFR-TYPE CYSTEINE-RICH REGION.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M35027; AAA48188.1; -
DR PIR: A42523; A42523.
DR PROSITE: PS00652; TNFR_NGFR_1; 1.
DR PROSITE: PS50050; TNFR_NGFR_2; 1.
DR PFAM: PF00020; TNFR_c6; 1.
KW Late protein; Repeat.
FT DOMAIN 36 103 2 X TNFR-CYS.
FT REPEAT 36 73 TNFR-CYS 1.
FT REPEAT 74 103 TNFR-CYS 2 (INCOMPLETE).
SQ SEQUENCE 103 AA; 12032 MW; 394241D8 CRC32;

Query Match 88.9%; Score 40; DB 1; Length 103;
```



CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: AE000321; AAC75381.1; -

DR EMBL: D90863; CAB22095.1; -

DR EMBL: D79661; AAB36529.1; -

DR EMBL: M29962; CAB25573.1; -

DR EMBL: M15541; -; NOT\_ANNOTATED\_CDS.

DR EMBL: M15541; -; NOT\_ANNOTATED\_CDS.

DR EMBL: M15541; -; NOT\_ANNOTATED\_CDS.

DR EMBL: M15541; -; NOT\_ANNOTATED\_CDS.

DR EMBL: M15541; -; NOT\_ANNOTATED\_CDS.

DR EMBL: M15541; -; NOT\_ANNOTATED\_CDS.

DR EMBL: M15541; -; NOT\_ANNOTATED\_CDS.

DR EMBL: M15541; -; NOT\_ANNOTATED\_CDS.

DR EMBL: M15541; -; NOT\_ANNOTATED\_CDS.

DR EMBL: M15541; -; NOT\_ANNOTATED\_CDS.

DR EMBL: M15541; -; NOT\_ANNOTATED\_CDS.

DR EMBL: M15541; -; NOT\_ANNOTATED\_CDS.

DR EMBL: M15541; -; NOT\_ANNOTATED\_CDS.

DR EMBL: M15541; -; NOT\_ANNOTATED\_CDS.

DR EMBL: M15541; -; NOT\_ANNOTATED\_CDS.

DR EMBL: M15541; -; NOT\_ANNOTATED\_CDS.

DR EMBL: M15541; -; NOT\_ANNOTATED\_CDS.

DR EMBL: M15541; -; NOT\_ANNOTATED\_CDS.

DR EMBL: M15541; -; NOT\_ANNOTATED\_CDS.

DR EMBL: M15541; -; NOT\_ANNOTATED\_CDS.

DR EMBL: M15541; -; NOT\_ANNOTATED\_CDS.

DR EMBL: M15541; -; NOT\_ANNOTATED\_CDS.

DR EMBL: M15541; -; NOT\_ANNOTATED\_CDS.

DR EMBL: M15541; -; NOT\_ANNOTATED\_CDS.

DR EMBL: M15541; -; NOT\_ANNOTATED\_CDS.

DR EMBL: M15541; -; NOT\_ANNOTATED\_CDS.

DR EMBL: M15541; -; NOT\_ANNOTATED\_CDS.

DR EMBL: M15541; -; NOT\_ANNOTATED\_CDS.

DR EMBL: M15541; -; NOT\_ANNOTATED\_CDS.

DR EMBL: M15541; -; NOT\_ANNOTATED\_CDS.

DR EMBL: M15541; -; NOT\_ANNOTATED\_CDS.

DR EMBL: M15541; -; NOT\_ANNOTATED\_CDS.

DR EMBL: M15541; -; NOT\_ANNOTATED\_CDS.

DR EMBL: M15541; -; NOT\_ANNOTATED\_CDS.

DR EMBL: M15541; -; NOT\_ANNOTATED\_CDS.

DR EMBL: M15541; -; NOT\_ANNOTATED\_CDS.

DR EMBL: M15541; -; NOT\_ANNOTATED\_CDS.

DR EMBL: M15541; -; NOT\_ANNOTATED\_CDS.

DR EMBL: M15541; -; NOT\_ANNOTATED\_CDS.

DR EMBL: M15541; -; NOT\_ANNOTATED\_CDS.

DR EMBL: M15541; -; NOT\_ANNOTATED\_CDS.

DR EMBL: M15541; -; NOT\_ANNOTATED\_CDS.

DR EMBL: M15541; -; NOT\_ANNOTATED\_CDS.

DR EMBL: M15541; -; NOT\_ANNOTATED\_CDS.

DR EMBL: M15541; -; NOT\_ANNOTATED\_CDS.

DR EMBL: M15541; -; NOT\_ANNOTATED\_CDS.

DR EMBL: M15541; -; NOT\_ANNOTATED\_CDS.

DR EMBL: M15541; -; NOT\_ANNOTATED\_CDS.

DR EMBL: M15541; -; NOT\_ANNOTATED\_CDS.

DR EMBL: M15541; -; NOT\_ANNOTATED\_CDS.

DR EMBL: M15541; -; NOT\_ANNOTATED\_CDS.

DR EMBL: M15541; -; NOT\_ANNOTATED\_CDS.

DR EMBL: M15541; -; NOT\_ANNOTATED\_CDS.

DR EMBL: M15541; -; NOT\_ANNOTATED\_CDS.

DR EMBL: M15541; -; NOT\_ANNOTATED\_CDS.

DR EMBL: M15541; -; NOT\_ANNOTATED\_CDS.

DR EMBL: M15541; -; NOT\_ANNOTATED\_CDS.

DR EMBL: M15541; -; NOT\_ANNOTATED\_CDS.

DR EMBL: M15541; -; NOT\_ANNOTATED\_CDS.

Matches 5; Conservative 1; Mismatches 7; Indels 0; Gaps 0;  
Db 136 PRNCNRETSVILM 148  
QY 2 PXXXXXXAVILM 14

RESULT 3

ID YDIJ\_BACSU STANDARD; PRT; 254 AA.

AC O05523;

DT 01-NOV-1997 (Rel. 35; Created)

DT 01-NOV-1997 (Rel. 35; Last sequence update)

DT 15-JUL-1998 (Rel. 36; Last annotation update)

DE HYPOTHETICAL 29.1 KD PROTEIN IN PROB-GROES INTERGENIC REGION.

GN YDIJ.

OS Bacillus subtilis.

OC Bacteria; Firmicutes; Bacillus/Clostridium group;

OC Bacillus/staphylococcus group; Bacillus.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=168;

RA SADAIE Y., YATA K., FUJITA M., SAGAI H., ITAYA M., KASAHARA Y.,

OGASAWARA N.

RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.

CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).

CC -!- SIMILARITY: BELONGS TO THE UPF0032 FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

CC EMBL: D88802; BAAL9723.1; -

CC EMBL: D88802; BAAL9723.1; -

CC SUBTILIST; BGI2207; YDIJ.

CC PROSITE; PS01218; UPF0032; 1.

CC PFAM; PF00902; UPF0032; 1.

CC KW Hypothetical protein; Transmembrane.

CC FT TRANSMEM 24 44 POTENTIAL.

CC FT TRANSMEM 67 87 POTENTIAL.

CC FT TRANSMEM 112 132 POTENTIAL.

CC FT TRANSMEM 157 177 POTENTIAL.

CC FT TRANSMEM 187 207 POTENTIAL.

CC FT TRANSMEM 212 232 POTENTIAL.

CC SQ SEQUENCE 254 AA; 29056 MW; 1EC13F83 CRC32;

Query Match 93.3%; Score 42; DB 1; Length 254;

Best Local Similarity 38.5%; Pred. No. 6.12e+00;

Matches 5; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

Db 164 PFGLLFQMPVILM 176

QY 2 PXXXXXXAVILM 14

RESULT 4

ID SCTL\_YEAST STANDARD; PRT; 759 AA.

AC P32784; Q07062;

DT 01-OCT-1993 (Rel. 27; Created)

DT 01-NOV-1997 (Rel. 35; Last sequence update)

DT 15-JUL-1998 (Rel. 36; Last annotation update)

DE CTRL SUPPRESSOR PROTEIN.

GN SCTL1 OR YBL011W OR YBL0315 OR YBL0309

OS Saccharomyces cerevisiae (Baker's yeast).

OC Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;

OC Saccharomycetaceae; Saccharomyces.

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE: 95332274.

RA MATSUSHITA M., NIKAWA J.;

\*\*\*\*\*  
W P S R L  
\*\*\*\*\*  
(TM)  
\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Sat Apr 15 01:27:40 2000: MasPar time 3.26 Seconds  
Tabular output not generated. 128.155 Million cell updates/sec

Title: >US-08-452-843-23  
Description: (1-14) from US08452843.pep  
Perfect Score: 45  
Sequence: 1 XPXXXXXXXVILM 14

Scoring table: PAM 150  
Gap 15

Searched: 82229 seqs, 29864866 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: swiss-prot38  
1:swissprot

Statistics: Mean 20.915; Variance 19.582; scale 1.068

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	43	95.6	331	1	DIV_ECOLI	3.41e+00
2	43	95.6	382	1	Y022_NPVAC	3.41e+00
3	42	93.3	254	1	Y01J_BACSU	6.12e+00
4	42	93.3	759	1	SC11_YEAST	6.12e+00
5	41	91.1	382	1	Y022_NPVOP	1.09e+01
6	40	88.9	103	1	VA53_VACCC	1.90e+01
7	40	88.9	103	1	VA53_VACCC	1.90e+01
8	39	86.7	142	1	Y1C9_YEAST	3.30e+01
9	39	86.7	229	1	RPE_CHLPN	3.30e+01
10	39	86.7	263	1	H225_MOUSE	3.30e+01
11	39	86.7	263	1	H225_MOUSE	3.30e+01
12	39	86.7	263	1	H22B_RAT	3.30e+01
13	39	86.7	265	1	H22A_MOUSE	3.30e+01
14	39	86.7	265	1	H22A_MOUSE	3.30e+01
15	39	86.7	265	1	H22D_MOUSE	3.30e+01
16	39	86.7	314	1	IUNH_CRIPA	3.30e+01
17	39	86.7	354	1	ACOD_MESAU	3.30e+01
18	39	86.7	355	1	ACOD_MOUSE	3.30e+01
19	39	86.7	358	1	ACOD_MOUSE	3.30e+01
20	39	86.7	358	1	ACOD_MOUSE	3.30e+01
21	39	86.7	359	1	ACOD_HUMAN	3.30e+01
22	39	86.7	573	1	Y1B1_BACSU	3.30e+01
23	39	86.7	2238	1	RRPL_BUNYW	3.30e+01

ID	DIV	ECOLI	STANDARD	PRT	331	AA
AC	P15286	P77706				
DT	01-APR-1990	(Rel. 14, Created)				
DT	01-NOV-1997	(Rel. 35, Last sequence update)				
DT	01-NOV-1997	(Rel. 35, Last annotation update)				
DE	DIV	PROTEIN				
GN	Div.					
OS	Escherichia coli.					
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;					
OC	Escherichia.					
RC	[1]					
RP	SEQUENCE FROM N.A.					
RC	STRAIN-K12 / MG1655;					
RC	MEDLINE; 97426617					
RA	BATTNER F.R., PLUNKETT G. III, BLOCH C.A., PERNA N.T., BURLAND V.,					
RA	RILEY M., COLLADO-VIDES J., GLASNER F.D., RODE C.K., MAYHEW G.F.,					
RA	GREGOR J., DAVIS N.W., KIRKPATRICK H.A., GOEDEN M.A., ROSE D.J.,					
RA	MAU B., SHAO Y.;					
RT	"The complete genome sequence of Escherichia coli K-12.";					
RL	Science 277:1453-1474(1997).					
RN	[2]					
RP	SEQUENCE FROM N.A.					
RC	STRAIN-K12;					
RA	ALBA H., BABA T., FUJITA K., HAYASHI K., HONJO A., HORIUCHI T.,					
RA	IKEMOTO K., INADA T., ISONO K., ITOH T., KANAI K., KASAI H.,					
RA	KASHIMOTO K., KIM S., KIMURA S., KITAGAWA M., KITAKAWA M., MAKINO K.,					
RA	MASUDA S., MIKI T., MIZOBUCHI K., MORI H., MOTOMURA K., NAKAMURA Y.,					
RA	NASHIMOTO H., NISHIO Y., OSHIMA T., SAITO N., SAMPEI G., SEKI Y.,					
RA	TAGAMI H., TAKEMOTO K., WADA C., YAMAMOTO Y., YANO M.;					
RL	Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.					
RN	[3]					
RP	SEQUENCE FROM N.A.					
RC	STRAIN-K12 / W3110;					
RA	PEASE A.J., SCHOENLEIN P.V., WINKLER M.E.;					
RL	Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.					
RN	[4]					
RP	SEQUENCE OF 1-39 FROM N.A.					
RC	STRAIN-K12;					
RC	MEDLINE; 90036695.					
RA	SCHOENLEIN P.V., ROA B.B., WINKLER M.E.;					
RT	"Divergent transcription of pdxB and homology between the pdxB and					
RT	serA gene products in Escherichia coli K-12.";					
RL	J. Bacteriol. 171:6084-6092(1989).					
CC	-----					
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration					
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation					



RL Brain Res. Mol. Brain Res. 45:90-98(1997).

DR EMBL; D84477; BAA20863.1; -.

DR HSSP; P06749; 1A2B.

DR PFAM; PF00071; ras; 1.

FT NON\_TER 1 1

FT NON\_TER 174 174

SQ SEQUENCE 174 AA; 19686 MW; 064801DE CRC32;

Query Match

91.1%; Score 41; DB 11; Length 174;

Best Local Similarity 41.7%; Pred. No. 3.01e+01;

Matches 5; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Db 61 PLSYPDPTDVILM 72

| | | |

QY 2 PXXXXXXAVILM 13

RESULT 15

ID O93469 PRELIMINARY; PRT; 174 AA.

AC O93469;

DT 01-NOV-1998 (TrEMBLrel. 08, Created)

DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)

DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)

DE GTPASE CRHOC (FRAGMENT).

GN CRHOC.

OS Gallus gallus (Chicken).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Archosauria; Aves;

OC Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=NEURAL RETINA;

RX MEDLINE; 97402517.

RA MALOSIO M.L., GILARDELLI D., PARIS S., ALBERTINAZZI C., DE CURTIS I.;

RT "Differential expression of distinct members of Rho family GTP-binding

RT proteins during neuronal development: identification of Rac1B, a new

RT neural-specific member of the family.";

RL J. Neurosci. 17:6717-6728(1997).

DR EMBL; U79759; AAC18964.1; -.

DR HSSP; P06749; 1A2B.

DR PFAM; PF00071; ras; 1.

FT NON\_TER 1 1

SQ SEQUENCE 174 AA; 20032 MW; 651ADCE5 CRC32;

Query Match

91.1%; Score 41; DB 13; Length 174;

Best Local Similarity 41.7%; Pred. No. 3.01e+01;

Matches 5; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Db 52 PLSYPDPTDVILM 63

| | | |

QY 2 PXXXXXXAVILM 13

Search completed: Sat Apr 15 01:22:57 2000

Job time : 94 secs.

OC Penaeus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-BRANCHIA;  
RA GENDREAU S., LEE R., MIALHE E.;  
RL Submitted (FEB-1994) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE-BRANCHIA;  
RA GENDREAU S.;  
RL Thesis (1992), UMR 9947, IFREMER-CNRS-University of Montpellier II, France.  
DR EMBL; 230080; CA82898.1; -.  
DR HSSP; P06749; 1A2B.  
SQ SEQUENCE 34 AA; 3967 MW; D81847BA CRC32;  
  
Query Match 91.1%; Score 41; DB 5; Length 34;  
Best Local Similarity 41.7%; Pred. No. 3.01e+01;  
Matches 5; Conservative 0; Mismatches 7; Indels 0; Gaps 0;  
  
Db 5 PLSYPDVTILM 16  
QY 2 PXXXXXXAVILM 13  
  
RESULT 11  
ID P79275 PRELIMINARY; PRT; 66 AA.  
AC P79275;  
DT 01-MAY-1997 (TREMBLrel. 03, Created)  
DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)  
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)  
DE RHO A (FRAGMENT).  
OS Sus scrofa (Pig).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-DOMESTIC; TISSUE-PULMONARY VEIN;  
RA MEDLINE; 97040692.  
RA NISHIMURA J., SAKITHARA C., ZHOU Y., KANAIDE H.;  
RT "Expression of rho A and rho kinase mRNAs in porcine vascular smooth muscle.";  
RL Biochem. Biophys. Res. Commun. 227:750-754(1996).  
DR EMBL; D89492; BAA13966.1; -.  
DR HSSP; P06749; 1A2B.  
FT NON\_TER 1 1  
FT NON\_TER 66 66  
SQ SEQUENCE 66 AA; 7562 MW; D2668D6A CRC32;  
  
Query Match 91.1%; Score 41; DB 6; Length 66;  
Best Local Similarity 41.7%; Pred. No. 3.01e+01;  
Matches 5; Conservative 0; Mismatches 7; Indels 0; Gaps 0;  
  
Db 36 PLSYPDVTILM 47  
QY 2 PXXXXXXAVILM 13  
  
RESULT 12  
ID Q9X0H0 PRELIMINARY; PRT; 113 AA.  
AC Q9X0H0;  
DT 01-NOV-1999 (TREMBLrel. 12, Created)  
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)  
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)  
DE ANTI-SIGMA FACTOR ANTAGONIST, PUTATIVE.  
GN TM1081.  
OS Thermotoga maritima.  
OC Bacteria; Thermotogales; Thermotoga.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 99287316.  
RA NELSON K.E., CLAYTON R.A., GILL S.R., GWINN M.L., DODSON R.J.,  
RA HAFT D.H., HICKEY E.K., PETERSON J.D., NELSON W.C., KETCHUM K.A.,

RA MCDONALD L., UTTERBACK T.R., MALEK J.A., LINHER K.D., GARRETT M.M.,  
RA STEWART A.M., COTTON M.D., PRATT M.S., PHILLIPS C.A., RICHARDSON D.,  
RA HEIDELBERG J., SUTTON G.G., FLEISCHMANN R.D., WHITE O., SALZBERG S.L.,  
RA SMITH H.O., VENTER J.C., FRASER C.M.;  
RT "Evidence for lateral gene transfer between Archaea and bacteria from genome sequence of Thermotoga maritima.";  
RL Nature 399:323-329(1999).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA NELSON K.E., CLAYTON R.A., GILL S.R., GWINN M.L., DODSON R.J.,  
RA HAFT D.H., HICKEY E.K., PETERSON J.D., NELSON W.C., KETCHUM K.A.,  
RA MCDONALD L., UTTERBACK T.R., MALEK J.A., LINHER K.D., GARRETT M.M.,  
RA STEWART A.M., COTTON M.D., PRATT M.S., PHILLIPS C.A., RICHARDSON D.,  
RA HEIDELBERG J., SUTTON G.G., FLEISCHMANN R.D., WHITE O., SALZBERG S.L.,  
RA SMITH H.O., VENTER J.C., FRASER C.M.;  
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AE001768; AAD36158.1; -.  
SQ SEQUENCE 113 AA; 12873 MW; 0BF7FB73 CRC32;  
  
Query Match 91.1%; Score 41; DB 2; Length 113;  
Best Local Similarity 41.7%; Pred. No. 3.01e+01;  
Matches 5; Conservative 0; Mismatches 7; Indels 0; Gaps 0;  
  
Db 3 PYKIVDDVVILM 14  
QY 2 PXXXXXXAVILM 13  
  
RESULT 13  
ID O86949 PRELIMINARY; PRT; 113 AA.  
AC O86949;  
DT 01-NOV-1998 (TREMBLrel. 08, Created)  
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)  
DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)  
DE HYPOTHETICAL 12.9 KD PROTEIN.  
OS Thermotoga neapolitana.  
OC Bacteria; Thermotogales; Thermotoga.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-22706-MC24;  
RA ZVERLOV V.V.;  
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AJ007446; CAA07514.1; -.  
KW Hypothetical protein.  
SQ SEQUENCE 113 AA; 12862 MW; 1A94BD59 CRC32;  
  
Query Match 91.1%; Score 41; DB 2; Length 113;  
Best Local Similarity 41.7%; Pred. No. 3.01e+01;  
Matches 5; Conservative 0; Mismatches 7; Indels 0; Gaps 0;  
  
Db 3 PYKIVEDVVILM 14  
QY 2 PXXXXXXAVILM 13  
  
RESULT 14  
ID O35791 PRELIMINARY; PRT; 174 AA.  
AC O35791;  
DT 01-JAN-1998 (TREMBLrel. 05, Created)  
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)  
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)  
DE RHOA (FRAGMENT).  
GN RHOA.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 97259573.  
RA YOSHIMURA S., SAKAI H., NAKASHIMA S., NOZAWA Y., SHINODA J., SAKAI N.,  
RA YAMADA H.;  
RT "Differential expression of Rho family GTP-binding proteins and protein kinase C isozymes during C6 glial cell differentiation.";

```
RC STRAIN=H37RV;
RX MEDLINE; 96181548.
RA PHILIPP W.J., POULET S., EIGLMEIER K., PASCOPELLA L.,
RA BALASUBRAMANIAN V., HEYM B., BERGH S., BLOOM B.R., JACOBS W.R. JR.,
RA COLE S.T.;
RT "An integrated map of the genome of the tubercle bacillus,
RT Mycobacterium tuberculosis H37RV, and comparison with Mycobacterium
RT leprae.";
RL Proc. Natl. Acad. Sci. U.S.A. 93:3132-3137(1996).
DR EMBL; ALQ08967; CAA1544.1; -.
DR PFAM; PF01580; FtsK_SpoIIIE; 1.
KW Hypothetical protein.
SQ SEQUENCE 883 AA; 94405 MW; FADC8E28 CRC32;

Query Match 95.6%; Score 43; DB 2; Length 883;
Best Local Similarity 41.7%; Pred. No. 9.95e+00;
Matches 5; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

Db 193 PLVAAAVAVVLM 204
QY 2 PXXXXXXAVILM 13

RESULT 7
ID O05560 PRELIMINARY; PRT; 886 AA.
AC O05560;
DT 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE HYPOTHELICAL 94.9 KD PROTEIN.
GN MLCB33.09C.
OS Mycobacterium leprae.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
RN [1]
RP SEQUENCE FROM N.A.
RA BADCOCK K., CHURCHER C.M.;
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA BARRELL B.G., RAJANDRAM M.A.;
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE; 93188700.
RA EIGLMEIER K., HONORE N., WOODS S.A., CAUDRON B., COLE S.T.;
RT "Use of an ordered cosmid library to deduce the genomic organization
RT of Mycobacterium leprae.";
RL Mol. Microbiol. 7:197-206(1993).
DR EMBL; Z94723; CAB08120.1; -.
DR PFAM; PF01580; FtsK_SpoIIIE; 1.
KW Hypothetical protein.
SQ SEQUENCE 886 AA; 94895 MW; 853C5531 CRC32;

Query Match 95.6%; Score 43; DB 2; Length 886;
Best Local Similarity 41.7%; Pred. No. 9.95e+00;
Matches 5; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

Db 181 PLVIAAVAVVLM 192
QY 2 PXXXXXXAVILM 13

RESULT 8
ID Q91778 PRELIMINARY; PRT; 1087 AA.
AC Q91778;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE INTEGRIN ALPHA 6 SUBUNIT (FRAGMENT).
GN INTA6.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Amphibia;
```

```
OC Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae; Xenopodinae;
OC Xenopus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WILD TYPE;
RX MEDLINE; 9631932.
RA LALLIER T.E., WHITTAKER C.A., DESIMONE D.W.;
RT "Integrin alpha 6 expression is required for early nervous system
RT development in xenopus laevis.";
RL Development 122:2539-2554(1996).
DR EMBL; L35051; AAB4884.1; -.
DR PROSITE; PS00242; INTEGRIN_ALPHA; 1.
DR PFAM; PF00357; Integrin_A; 3.
KW Integrin.
RN [1]
RP NON_TER
SQ SEQUENCE 1087 AA; 121720 MW; F3A1E50A CRC32;

Query Match 95.6%; Score 43; DB 13; Length 1087;
Best Local Similarity 41.7%; Pred. No. 9.95e+00;
Matches 5; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

Db 289 PRANHSAGVAVLM 300
QY 2 PXXXXXXAVILM 13

RESULT 9
ID Q26078 PRELIMINARY; PRT; 34 AA.
AC Q26078;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE RAS-HOMOLOGUE (PUTATIVE) (FRAGMENT).
OS Penaeus monodon (Pencoid shrimp).
OC Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
OC Eumalacostraca; Eucarida; Decapoda; Dendrobranchiata; Penaeidae;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-BRANCHIA;
RA GENDREAU S., LEE R., MIALHE E.;
RL Submitted (FEB-1994) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE-BRANCHIA;
RA GENDREAU S.;
RL Thesis (1992), UMR 9947, IFREMER-CNRS-University of Montpellier II,
RL France.
DR EMBL; Z30081; CAA82899.1; -.
DR HSSP; P06749; 1A2B.
FT NON_TER 1 1
FT NON_TER 34 34
SQ SEQUENCE 34 AA; 3953 MW; 2A98F7F4 CRC32;

Query Match 91.1%; Score 41; DB 5; Length 34;
Best Local Similarity 41.7%; Pred. No. 3.01e+01;
Matches 5; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Db 5 PLSYPDTDVILM 16
QY 2 PXXXXXXAVILM 13

RESULT 10
ID Q26077 PRELIMINARY; PRT; 34 AA.
AC Q26077;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE PUTATIVE PARTIAL RAS-HOMOLOGOUS PRECURSOR.
OS Penaeus monodon (Pencoid shrimp).
OC Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
OC Eumalacostraca; Eucarida; Decapoda; Dendrobranchiata; Penaeidae;
```

```

DE ACID PHOSPHATASE TYPE 5.
OS Emeritella nidulans (Aspergillus nidulans).
OC Eukaryota; Fungi; Ascomycota; Euascomycetes; Plectomycetes;
OC Eurotiales; Trichocomaceae; Emeritella.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FGSC4; TISSUE=MYCELIUM;
RX MACRAE W.D., BUXTON F.P., SIBLEY S., GARVEN S., GWYNNE D., ARST H.N.,
RA DAVIES R.W.;
RT "Characterization of an Aspergillus nidulans genomic DNA fragment
RT conferring phosphate-non-repressible acid-phosphatase activity.";
RL Gene 130:247-251(1993).
DR EMBL; M96993; AAA33288.1; -.
SQ SEQUENCE 113 AA; 12638 MW; FAF76117 CRC32;

Query Match 95.6%; Score 43; DB 3; Length 113;
Best Local Similarity 41.7%; Pred. No. 9.95e+00;
Matches 5; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

Db 57 PIVAEHLTVILM 68
QY 2 PXXXXXXAVILM 13
|
|:|
|:|

RESULT 3
ID Q92T87 PRELIMINARY; PRT; 378 AA.
AC Q92T87;
DT 01-MAY-1999 (TRENBLrel. 10, Created)
DT 01-MAY-1999 (TRENBLrel. 10, Last sequence update)
DT 01-MAY-1999 (TRENBLrel. 10, Last annotation update)
DE PELOTA.
GN PELI.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
OC Arabidopsis.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV, COLUMBIA;
RA CARYL A.P., FRANKLIN F.C.H., JONES G.;
RT "cDNA sequence of the Arabidopsis thaliana gene AtPelot1.";
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF069497; AAC82379.1; -.
SQ SEQUENCE 378 AA; 42519 MW; C1FD2692 CRC32;

Query Match 95.6%; Score 43; DB 10; Length 378;
Best Local Similarity 41.7%; Pred. No. 9.95e+00;
Matches 5; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

Db 131 PAASADLAVILM 142
QY 2 PXXXXXXAVILM 13
|
|:|
|:|

RESULT 4
ID O59101 PRELIMINARY; PRT; 510 AA.
AC O59101;
DT 01-AUG-1998 (TRENBLrel. 07, Created)
DT 01-AUG-1998 (TRENBLrel. 07, Last sequence update)
DT 01-MAY-1999 (TRENBLrel. 10, Last annotation update)
DE 510AA LONG HYPOTHETICAL PROTEIN.
GN PH1431.
OS Pyrococcus horikoshii.
OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OT3;
RX MEDLINE; 98344137.
RA KAWARABAYASI Y., SAWADA M., HORIKAWA H., HAIKAWA Y., HINO Y.,
RA YAMAMOTO S., SEKINE M., BABA S., KOSUGI H., HOSOTAMA A., NAGAI Y.,
RA SAKAI M., OGURA K., OTUKA K., NAKAZAWA H., TAKAMIYA M., OHFUKU Y.,

```

```

RA FUNAHASHI T., TANAKA T., KUDOH Y., YAMAZAKI J., KUSHIDA N., OGUCHI A.,
RA AOKI K., NAKAMURA Y., ROBB T.F., HORIKOSHI K., MASUCHI Y., SHIZUYA H.,
RA KIKUCHI H.;
RT "Complete sequence and gene organization of the genome of a hyper-
RT thermophilic archaeobacterium, Pyrococcus horikoshii OT3.";
RL DNA Res 5:55-76(1998).
DR EMBL; AF000006; BAA30538.1; -.
DR PFAM; PF00361; oxidored_51; 1.
SQ SEQUENCE 510 AA; 55287 MW; 1F9C3173 CRC32;

Query Match 95.6%; Score 43; DB 1; Length 510;
Best Local Similarity 41.7%; Pred. No. 9.95e+00;
Matches 5; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

Db 466 PMFILVIAIILM 477
QY 2 PXXXXXXAVILM 13
|
|:|
|:|

RESULT 5
ID Q90518 PRELIMINARY; PRT; 530 AA.
AC Q90518;
DT 01-NOV-1996 (TRENBLrel. 01, Created)
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
DT 01-JAN-1999 (TRENBLrel. 09, Last annotation update)
DE SODIUM CHANNEL ALPHA SUBUNIT (FRAGMENT).
OS Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Actinopterygii;
OC Neopterygii; Teleostei; Euteleostei; Acanthopterygii; Percomorpha;
OC Tetraodontiformes; Tetraodontidae; Tetraodontidae; Fugu.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RA NAKAZAWA A.;
RL Submitted (AUG-1994) to the EMBL/GenBank/DBJ databases.
DR EMBL; D37976; BAA07194.1; -.
DR PFAM; PF00520; Ion_Trans; 1.
KW Ionic channel.
FT NON_TER 530
SQ SEQUENCE 530 AA; 60164 MW; 1D3E779B CRC32;

Query Match 95.6%; Score 43; DB 13; Length 530;
Best Local Similarity 41.7%; Pred. No. 9.95e+00;
Matches 5; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

Db 190 PNWLDFSVILM 201
QY 2 PXXXXXXAVILM 13
|
|:|
|:|

RESULT 6
ID Q33290 PRELIMINARY; PRT; 883 AA.
AC Q33290;
DT 01-JAN-1998 (TRENBLrel. 05, Created)
DT 01-JAN-1998 (TRENBLrel. 05, Last sequence update)
DT 01-NOV-1999 (TRENBLrel. 12, Last annotation update)
DE HYPOTHETICAL 94.4 KD PROTEIN.
GN MTV002.13C.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RA MURPHY L., HARRIS D.;
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RA PARKHILL J., BARRELL B.G., RAJANDREAM M.A.;
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.

```

\*\*\*\*\*

WQSRH  
(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Sat Apr 15 01:21:23 2000; MasPar time 7.36 Seconds  
Tabular output not generated.  
122.514 Million cell updates/sec

Title: >US-08-452-843-22  
Description: (1-13) from US08452843.pgp  
Perfect Score: 45  
Sequence: 1 XPXXXXXAVILM 13

Scoring table: PAM 150  
Gap 15

Searched: 225878 seqs, 69334122 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: sptrembl12  
1:sp-archaea 2:sp-bacteria 3:sp-fungi 4:sp-human  
5:sp-invertebrate 6:sp-mammal 7:sp-mhc 8:sp-organelle  
9:sp-phage 10:sp-plant 11:sp-rodent 12:sp-unclassified  
13:sp-vertebrate 14:sp-virus

Statistics: Mean 20.252; Variance 20.472; scale 0.989

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query % Match	Length	ID	Description	Pred. No.
1	45	100.0	350	2 P77230	MITOCHONDRIAL TRIFUNCT	3.17e+00
2	43	95.6	113	3 Q00745	ACID PHOSPHATASE TYPE	9.95e+00
3	43	95.6	378	10 Q92T87	PELOTA.	9.95e+00
4	43	95.6	510	1 Q59101	510AA LONG HYPOTHETICA	9.95e+00
5	43	95.6	530	13 Q90518	SODIUM CHANNEL ALPHA S	9.95e+00
6	43	95.6	883	2 Q33290	HYPOTHETICAL 94.4 KD P	9.95e+00
7	43	95.6	1087	13 Q91778	INTEGRIN ALPHA 6 SUBUN	9.95e+00
8	43	95.6	34	5 Q26078	RAS-HOMOLOGUE (PUTATIV	3.01e+01
9	41	91.1	34	5 Q26077	PUTATIVE PARTIAL RAS-H	3.01e+01
10	41	91.1	66	6 P79275	RHO A (FRAGMENT)	3.01e+01
11	41	91.1	113	2 Q9X0H0	ANTI-SIGMA FACTOR ANTA	3.01e+01
12	41	91.1	113	2 Q86949	HYPOTHETICAL 12.9 KD P	3.01e+01
13	41	91.1	174	11 Q35791	RHOA (FRAGMENT)	3.01e+01
14	41	91.1	174	13 Q93469	GTPASE CRHOC (FRAGMENT	3.01e+01
15	41	91.1	192	5 Q97041	RHO1 GTPASE	3.01e+01
16	41	91.1	193	11 Q88336	RHO FAMILY GTPASE	3.01e+01
17	41	91.1	193	13 Q9W760	SMALL RHO-LIKE GTPASE	3.01e+01
18	41	91.1	193	13 Q93467	GTPASE CRHOC	3.01e+01
19	41	91.1	196	13 Q93468	GTPASE CRHOS	3.01e+01
20	41	91.1				

21	41	91.1	221	1 Q50523	HYPOTHETICAL 23.9 KD P	3.01e+01
22	41	91.1	292	13 Q92038	ACYL-COA DESATURASE (E	3.01e+01
23	41	91.1	353	2 Q92HE9	CHORISMATE SYNTHASE (E	3.01e+01
24	41	91.1	487	2 Q9X3X9	TETRACENOMYCIN C RESIS	3.01e+01
25	40	88.9	323	5 Q16197	D1065.5 PROTEIN.	5.16e+01
26	40	88.9	367	3 Q94642	PUTATIVE URACIL KINASE	5.16e+01
27	40	88.9	452	5 Q18924	SIMILAR TO ARYLISULFATA	5.16e+01
28	40	88.9	452	5 Q54193	HYPOTHETICAL 44.8 KD P	5.16e+01
29	40	88.9	637	5 Q44547	R02C2.1 PROTEIN.	5.16e+01
30	40	88.9	868	5 Q9Y1V3	TUNICATE RETINOIC ACID	5.16e+01
31	39	86.7	326	14 Q98044	OUTER CAPSID PROTEIN V	8.75e+01
32	39	86.7	326	14 Q98043	OUTER CAPSID PROTEIN V	8.75e+01
33	39	86.7	326	14 Q56347	OUTER CAPSID GLYCOPROT	8.75e+01
34	39	86.7	326	14 Q98046	OUTER CAPSID PROTEIN V	8.75e+01
35	39	86.7	326	14 Q98045	OUTER CAPSID PROTEIN V	8.75e+01
36	39	86.7	326	14 Q98047	OUTER CAPSID PROTEIN V	8.75e+01
37	39	86.7	326	14 Q98048	OUTER CAPSID PROTEIN V	8.75e+01
38	39	86.7	326	14 Q98037	OUTER CAPSID PROTEIN V	8.75e+01
39	39	86.7	326	14 Q86190	VP7.	8.75e+01
40	39	86.7	326	14 Q66782	VIRAL PROTEIN 7.	8.75e+01
41	39	86.7	326	14 Q66772	GLYCOPROTEIN VP7.	8.75e+01
42	39	86.7	326	14 Q86216	MAJOR OUTER CAPSID PRO	8.75e+01
43	39	86.7	326	14 P89071	GENE 9 (STRAIN IS2).	8.75e+01
44	39	86.7	400	1 Q9YD77	400AA LONG HYPOTHETICA	8.75e+01
45	39	86.7	1988	11 Q88421	VOLTAGE-GATED SODIUM C	8.75e+01

ALIGNMENTS

RESULT 1	PRELIMINARY;	PRT;	350 AA.
ID P77230;			
AC P77230;			
DT 01-FEB-1997 (Tremblrel. 02, Created)			
DT 01-FEB-1997 (Tremblrel. 02, Last sequence update)			
DT 01-NOV-1999 (Tremblrel. 12, Last annotation update)			
DE MITOCHONDRIAL TRIFUNCTIONAL ENZYME BETA SUBUNIT			
DE (CONTAINS: 3-KETOACYL-COA THIOLASE (EC 2.3.1.16)			
DE (ACETYL-COA ACYLTRANSFERASE) (BETA-KETOTHIOLASE)).			
GN HADHB.			
OS Escherichia coli.			
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;			
OC Escherichia.			
RN [1]			
RP SEQUENCE FROM N.A.			
RC STRAIN-K12;			
RA AIBA H., BABA T., FUJITA K., HAYASHI K., HONJO A., HORIUCHI T.,			
RA IKEMOTO K., INADA T., ISONO K., ISONO S., ITOH T., KANAI K., KASAI H.,			
RA KASHIMOTO K., KIM S., KIMURA S., KITAGAWA M., KITAKAWA M., MAKINO K.,			
RA MASUDA S., MIKI T., MIZOBUCHI K., MORI H., MOTOMURA K., NAKAMURA Y.,			
RA NASHIMOTO H., NISHIO Y., OSHIMA T., SAITO N., SAMPEI G., SEKI Y.,			
RA TAGAMI H., TAKEMOTO K., WADA C., YAMAMOTO Y., YANO M.,			
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.			
DR EMBL; D90864; BAA16197.1; -			
DR EMBL; D90865; BAA16203.1; -			
DR PFAM; PF00108; thiolase; 1.			
KW Transferase; Acyltransferase.			
SQ SEQUENCE 350 AA; 37998 MW; 22A06942 CRC32;			

Query Match 100.0%; Score 45; DB 2; Length 350;  
Best Local Similarity 50.0%; Pred. No. 3.17e+00;  
Matches 6; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 293 PLTDCGAAVILM 304

Qy 2 PXXXXXAVILM 13

RESULT 2 PRELIMINARY; PRT; 113 AA.

ID Q00745;  
AC Q00745;  
DT 01-NOV-1996 (Tremblrel. 01, Created)  
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)  
DT 01-NOV-1998 (Tremblrel. 08, Last annotation update)



**THIS PAGE BLANK (USPTO)**

CC polypeptides and identifying those with selective affinity for the  
 CC complex. Proteins containing WW domains are used for targeted drug  
 CC screening, i.e. to identify potential modulators of specific WW domain  
 CC interactions. The valency of the recognition unit is important in  
 CC determining specificity of interaction with WW domains. In multivalent  
 CC form specificity is relaxed, but not lost, so proteins containing WW  
 CC domains similar, but not identical, to the sequence of the peptides  
 CC target WW can be detected, including new polypeptides.  
 SQ Sequence 724 AA;

Query Match 88.9%; Score 40; DB 1; Length 724;  
 Best Local Similarity 38.5%; Pred. No. 3.06e+02;  
 Matches 5; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

Db 627 PVIQWFWKAVILM 639  
 QY 2 PXXXXXXVILM 14

Search completed: Sat Apr 15 01:26:44 2000  
 Job time : 35 secs.

Db 66 PFIVKTLIAMILM 78  
QY 2 PXXXXXXAVILM 14

RESULT 13  
ID R27737 standard; Protein: 416 AA.  
AC R27737;  
DE Sequence transcribed from third reading frame of  
DE Vaccinia virus DNA from positions 17201-18450.  
KW Virus vector; vaccinia virus; papillomavirus; HPV;  
KW immunotherapeutic; neutral site.  
OS Vaccinia virus.  
PN WO9216636-A.  
PD 01-OCT-1992.  
PF 10-MAR-1992; G00424.  
PR 14-MAR-1991; GB-005383.  
PI Bourns MEG, Inglis SC, Munro AJ;  
DR WPI: 92-349219/42.  
DR N-PSDB: Q29467.  
PT Recombinant virus vectors encoding human papillomavirus proteins  
PT - for treating and vaccinating against HPV infections and  
PT conditions caused by them, such as cervical cancer  
PS Disclosure: Fig 19; 83pp; English.  
CC To make a recombinant virus vector comprising human papillomavirus  
CC genes inserted into the vaccinia virus genome, neutral sites  
CC for insertion must be utilised such that replicative ability is not  
CC adversely affected. The neutral sites are identified by analysing  
CC the viral genome to identify ORFs which are likely to encode  
CC functional genes and selecting sites between such ORFs or within  
CC sequences for non-functional genes. The sequence shown is that  
CC transcribed from the vaccinia virus WR strain positions 17201-18450  
CC contg. the regions covered by the four fragments Salf, G, H and I.  
CC The sequence was transcribed in all three reading frames to determine  
CC genuine vaccinia virus genes via codon usage, thus determining neutral  
CC sites. HPV DNA sequences may be inserted neutral sites, e.g. those  
CC encoding E6 or E7 of HPV 16 and 18 or mutants of these proteins.  
CC The recombinant virus vector may be used immunotherapeutically to  
CC activate cells of the immune system against HPV.  
CC See also R27723-43.  
SQ Sequence 416 AA;

Query Match 88.9%; Score 40; DB 1; Length 416;  
Best Local Similarity 30.8%; Pred. No. 3.06e+02;  
Matches 4; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

Db 210 PPHYTTIPILM 222  
QY 2 PXXXXXXAVILM 14

RESULT 14  
ID W12375 standard; Protein: 450 AA.  
AC W12375;  
DE Human stromal cell NT3-gamma protein.  
DE Mouse; stromal cell; NT3; human; abnormal bone metabolism; growth; blood;  
KW nervous system functioning; glial stem cell.  
OS Homo sapiens.  
FH Key  
FT peptide 1..36  
FT /note= "signal peptide"  
FT protein 37..450  
FT /note= "mature protein"  
PN J08301898-A.  
PD 19-NOV-1996.  
PF 28-APR-1995; 128881.  
PR 28-APR-1995; JP-128881.  
PA (ONOY ) ONO PHARM CO LTD.  
DR WPI: 97-048321/05.

DR N-ESDB; T63204.  
PT DNA encoding polypeptide used to treat insufficient growth of  
PT blood-forming cells - and abnormal bone metabolism  
PS Claim 11: Page 23-24; 31pp; Japanese.  
CC This is the amino acid sequence of the human stromal cell protein  
CC NT3-gamma. The corresponding gene was isolated from cDNA library  
CC generated from polyA+ purified human glial stem cell line 198G. The  
CC probe was a fragment of the mouse NT3 gene (T63201). 80 positive clones  
CC were isolated. Restriction digestion and subsequent separation by gel  
CC electrophoresis revealed 3 major insert sizes: 1.9, 2.1 and 2.7 kb.  
CC Sequence analysis showed 3 new NT3 gene sequences designated NT3-alpha,  
CC (T63202) NT3-beta (T63203) and NT3-gamma. The proteins can be used to  
CC treat abnormal bone metabolism, irregular nervous system functioning or  
CC insufficient growth of blood forming cells.  
SQ Sequence 450 AA;

Query Match 88.9%; Score 40; DB 1; Length 450;  
Best Local Similarity 38.5%; Pred. No. 3.06e+02;  
Matches 5; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

Db 20 PCLWLLGAVILM 32  
QY 2 PXXXXXXAVILM 14

RESULT 15  
ID W36797 standard; Peptide: 724 AA.  
AC W36797;  
DE 23-APR-1998 (first entry)  
DE Novel human gene, designated WWP4.  
DE Peptide recognition unit; YAP WW domain binding protein; WBP-1; WBP-2;  
KW WW domain; cell signalling; growth regulation; cytoskeleton organisation;  
KW targeted drug screening; modulator; WW domain interaction; WWP4.  
OS Homo sapiens.  
FH Key  
FT Location/Qualifiers  
FT misc\_difference 1..3  
FT /note= "the nucleotides encoding these amino acids  
FT are not given in the specification"  
FT Domain 140..165  
FT /note= "claimed (claim 49) WW Domain 1"  
FT Domain 252..277  
FT /note= "claimed (claim 49) WW domain 3"  
FT Domain 303..328  
FT /note= "claimed (claim 49) WW domain 3"  
FT Domain 618..724  
FT /note= "claimed (claim 90) HECT domain"

PN W09737223-A1.  
PD 09-OCT-1997.  
PF 03-APR-1997; U05547.  
PR 03-APR-1996; US-630916.  
PA (CYTO-) CYTOGEN CORP.  
PA (UYNC-) UNIV NORTH CAROLINA.  
PI Fowlkes DM, Kay BK, Pirozzi G;  
DR WPI: 97-503234/46.  
DR N-PSDB: 195700.  
PT Identifying cell signalling and growth regulatory polypeptides by  
PT reaction with multivalent recognition complex - polypeptides are  
PT useful in targeted drug selection  
PS Claim 48: Fig 23; 220pp; English.  
CC The present sequence represents a novel protein WWP4. The WWP4 gene was  
CC identified and isolated from a cDNA expression library generated from  
CC LNCap prostate cancer cell line, using peptides W38063-64. These peptide  
CC recognition units are based on the sequences of WW domain binding domains  
CC of the alpha and gamma subunits of epithelial sodium channel protein.  
CC The WW domain is a small functional domain found in a large number of  
CC proteins from a variety of species including humans, nematodes and yeast.  
CC Its name is derived from the observation that two tryptophan residues,  
CC one in the amino terminal portion of the WW domain and one in the  
CC carboxyl terminal portion, are conserved. Most proteins containing WW  
CC domains have a function involving cell signalling and growth regulation  
CC or the organisation of the cytoskeleton. Polypeptides containing a WW  
CC domain are identified by treating a multivalent recognition unit complex  
CC that has selective binding affinity for a WW domain, with many

28-SEP-1998 (first entry)  
Human GDNF alpha-3 receptor protein #1.  
Glial cell line-derived neurotrophic factor alpha-3 receptor; GDNF;  
treatment; neurodegenerative disease; Parkinson's Disease; ALS; SMA;  
anyotrophic lateral sclerosis; spinal muscular atrophy; nerve; trauma;  
Huntington's Disease; Alzheimer's Disease; diabetic neuropathy; muscle;  
muscular dystrophy; diagnostic.  
OS Homo sapiens.  
FH Key Location/Qualifiers  
FT Protein 1..400  
FT /label= "GDNF alpha-3"  
FT /note= "Partial sequence"  
FT EP-846764-A2.  
PN 10-JUN-1998.  
PD 20-NOV-1997; 309375.  
PR 09-MAY-1997; GB-009463.  
PR 27-NOV-1996; GB-024677.  
PA (SMIK ) SMITHKLINE BEECHAM PLC.  
PI Lawrence GMP;  
DR N-PSDB; V35364.  
DR New factor alpha 3 receptor polypeptide and e.g. DNA and agonists -  
PT used to treat neuro degenerative diseases, muscular diseases and  
PT nerve and muscle trauma and in diagnostic assays  
PS Claim 4; Fig 2; 22pp; English.  
CC This sequence represents a novel glial cell line-derived neurotrophic  
CC factor alpha-3 receptor (GDNF alpha-3). This protein can be used to  
CC treat e.g. neurodegenerative diseases (such as Parkinson's Disease,  
CC anyotrophic lateral sclerosis (ALS), spinal muscular atrophy (SMA),  
CC Huntington's Disease, Alzheimer's Disease, diabetic neuropathy), and  
CC muscular diseases (including the muscular dystrophies) and nerve and  
CC muscle trauma and in diagnostic assays for such conditions.  
SQ Sequence 400 AA;

Query Match 88.9%; Score 40; DB 1; Length 400;  
Best Local Similarity 30.8%; Pred. No. 3.06e+02;  
Matches 4; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

Db 4 PLNRPRLPPVVL 16  
| :|||  
QY 2 PXXXXXXAVILM 14

RESULT 11  
ID W37463 standard; Protein; 400 AA.  
AC W37463;  
DT 21-MAY-1998 (first entry)  
DE Human Ret ligand RetL3.  
KW Ret ligand; RetL; RetL3; receptor; signal transduction; human;  
KW cell growth; renal cell; nerve cell; renal failure; nephritis;  
KW kidney transplant; toxic injury; hypoxic injury;  
KW neurodegeneration; motor neurone disease; multiple sclerosis;  
KW infection; meningitis; myelopathy; Creutzfeldt-Jakob disease;  
KW cranial nerve injury; spinal cord injury; Down's syndrome;  
KW cerebral palsy; Lyme disease; muscular dystrophy;  
KW myasthenia gravis; tumour; therapy.  
OS Homo sapiens.  
PN WO9744356-A2.  
PD 27-NOV-1997.  
PF 07-MAY-1997; U07726.  
PR 10-APR-1997; US-017427.  
PR 08-MAY-1996; US-017427.  
PR 07-JUN-1996; US-019300.  
PR 16-JUL-1996; US-021859.  
PA (BIOJ ) BIOGEN INC.  
PI Cate RL, Hession C, Sanicola-Nadel M;  
DR N-PSDB; V00251.  
DR New nucleic acid encoding ret receptor ligands and related proteins  
PT - vectors, transformed cells and antibodies, used for promoting cell  
PT growth and improving survival of injured cells, especially renal or  
PT nerve cells  
PS Claim 2; Page 85-86; 113pp; English.

CC This amino acid sequence comprises human Ret ligand (RetL) RetL3,  
CC deduced from cDNA clones (see V00251) isolated from a adult heart  
CC and spinal cord libraries. Rat and human RetL1, human RetL2 and  
CC mouse RetL3 sequences (see W37457-62) are also claimed. Human  
CC RetL3 is 34.3% identical to human RetL1, 34.9% identical to human  
CC RetL2 and 76.8% identical to murine RetL3. Ret ligand is a key  
CC component of the Ret signalling pathway that specifically  
CC interacts with Ret receptor protein, triggering Ret dimerisation  
CC and/or autophosphorylation of the Ret tyrosine kinase domain.  
CC Vectors containing RetL3 DNA and prokaryotic or eukaryotic host  
CC cells transformed or transfected with these vectors are claimed, as  
CC well as a method for production of RetL3, its soluble variants and  
CC fusion proteins with a toxin, imageable compound or radionuclide.  
CC RetL3, optionally when expressed from vectors in vivo, is used to  
CC promote growth of new tissue and survival of damaged tissue,  
CC particularly kidney or neural tissue. Typical applications are in  
CC renal failure, nephritis, kidney transplants, toxic or hypoxic  
CC injury, neurodegeneration, motor neurone disease, multiple sclerosis,  
CC bacterial, viral or prion infections (e.g. meningitis, myelopathy  
CC associated with HIV or Creutzfeldt-Jakob disease), cranial nerve or  
CC spinal cord injury, developmental disorders such as Down's syndrome  
CC and cerebral palsy, or conditions involving the peripheral nervous  
CC system (Lyme disease, muscular dystrophy and myasthenia gravis).  
CC Fusion proteins are used to deliver toxins etc. to Ret-expressing  
CC cells, especially tumours.  
SQ Sequence 400 AA;

Query Match 88.9%; Score 40; DB 1; Length 400;  
Best Local Similarity 30.8%; Pred. No. 3.06e+02;  
Matches 4; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

Db 4 PLNRPRLPPVVL 16  
| :|||  
QY 2 PXXXXXXAVILM 14

RESULT 12  
ID W70514 standard; Protein; 416 AA.  
AC W70514;  
DT 19-JAN-1999 (first entry)  
DE Candida glabrata IPC synthase.  
KW Candida; IPC synthase; fungus; inositolphosphoryl ceramide synthase;  
KW anti-fungal therapy; sphingolipid biosynthesis; phosphatidylinositol.  
OS Candida glabrata.  
PN EP-872485-A2.  
PD 21-OCT-1998.  
PF 14-APR-1998; 302866.  
PR 17-OCT-1997; US-082971.  
PR 15-APR-1997; US-043591.  
PR 22-APR-1997; US-044095.  
PR 13-MAY-1997; US-046348.  
PR 21-JUL-1997; US-053320.  
PA (ELL ) LILLY & CO ELL.  
PI Heider SA, Radding JA;  
DR WPI; 98-533879/46.  
DR N-PSDB; V33669.  
PT New inositolphosphoryl ceramide synthase genes from fungi - useful  
PT for identifying compounds for anti-fungal therapy  
PS Claim 2; Page 13-14; 53pp; English.  
CC The present sequence represents a pure inositolphosphoryl ceramide (IPC)  
CC synthase protein from a fungal cell, Candida glabrata. The present  
CC invention also describes a method for identifying inhibitory compounds  
CC of fungal IPC synthase protein activity. IPC synthase proteins are  
CC useful for identifying inhibitors of fungal sphingolipid biosynthesis,  
CC as the IPC synthase catalyses a step in the synthesis of  
CC inositolphosphoryl ceramide from ceramide and phosphatidylinositol.  
CC Fragments of IPC synthase proteins are also useful as probes or primers  
CC for identification and isolation of homologous genes.  
SQ Sequence 416 AA;

Query Match 88.9%; Score 40; DB 1; Length 416;  
Best Local Similarity 38.5%; Pred. No. 3.06e+02;  
Matches 5; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

PI Fox GM, Jing S, Wen D;  
 DR WPI: 99-080806/07.  
 DR N-PSDB: V99329.  
 PT New isolated glial cell line-derived neurotrophic factor receptors -  
 PT used to develop products for treating e.g. improperly functioning  
 PT dopaminergic nerve cells, Parkinson's disease, Alzheimer's disease  
 PT or amyotrophic lateral sclerosis  
 PS Claim 51; Fig 15; 31pp; English.  
 CC The present sequence represents a human glial cell-line derived  
 CC neurotrophic factor receptor (GDNFR)-related protein 3 (GRR3).  
 CC The protein has similar functions to GDNFR. GDNFR proteins are  
 CC functionally characterised by the ability to bind glial cell  
 CC line-derived neurotrophic factor (GDNF) and/or neurturin specifically,  
 CC and to act as part of a molecular complex which mediates or enhances  
 CC the signal transduction affects of GDNF and/or neurturin. The proteins  
 CC can be used for treating improperly functioning dopaminergic nerve  
 CC cells, Parkinson's disease, Alzheimer's disease or amyotrophic lateral  
 CC sclerosis. They can also be used for treating neurological disorders  
 CC associated with diabetes, glaucoma or other diseases and conditions  
 CC involving retinal ganglion cell degeneration, sensory neuropathy caused  
 CC by injury to, insults to, or degeneration of, sensory neurons,  
 CC pathological conditions, or disease or injury-related retinopathies.  
 CC The products can also be used for detection, diagnosis, drug screening  
 CC and gene therapy.  
 SQ Sequence 400 AA;

Query Match 88.9%; Score 40; DB 1; Length 400;  
 Best Local Similarity 30.8%; Pred. No. 3.06e+02;  
 Matches 4; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

Db 4 PLNRPRLPPVILM 16  
 | :|||  
 QY 2 PXXXXXXAVILM 14

RESULT 8  
 ID W84186 standard; Protein; 400 AA.  
 AC W84186;  
 DT 25-MAR-1999 (first entry)  
 DE Glial cell line-derived neurotrophic factor receptor gamma 2.  
 KW Glial cell line-derived neurotrophic factor; GDNF;  
 KW GDNFR-alpha; glial cell line-derived neurotrophic factor; GDNF;  
 KW neurodegenerative disease; amyotrophic lateral sclerosis; GDNFR-gamma2;  
 KW Parkinson's disease; schizophrenia; insomnia; tardive dyskinesia;  
 KW hypertension; pituitary adenomas; hyperprolactinemia; thyroid tumour;  
 KW renal disorder; kidney failure; gut dysfunction; regeneration;  
 KW cardiomyocyte; epithelium; hepatocyte.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT Peptide 1..31  
 FT /label= signal\_peptide  
 FT Protein 32..400  
 FT /label= mature\_protein  
 FT Domain 32..382  
 FT /note= "extracellular domain"  
 FT Domain 383..400  
 FT /note= "transmembrane domain"  
 FT WO9853069-A2.  
 PN 26-NOV-1998.  
 PD 20-MAY-1998; U10328.  
 PF 20-MAY-1997; US-884638.  
 PR 27-JUN-1997; US-047092.  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 PI Gentz RL, Hsu T, N1 J, Ruben SM, Young P;  
 DR WPI: 99-070150/06.  
 DR N-PSDB: V99334.  
 PT New isolated glial cell derived neurotrophic factor receptors - used  
 PT to develop products for treating e.g. neurodegenerative disorders,  
 PT schizophrenia, hypertension, tumours, renal disorders, kidney  
 PT failure or gut dysfunction  
 PS Claim 53; fig 7A-D; 156pp; English.  
 CC The present sequence represents a glial cell line-derived neurotrophic  
 CC factor receptor gamma 2 (GDNFR-gamma2). GDNFR-gamma2 shares high homology

CC with GDNFR-alpha, which is capable of complexing with glial cell  
 CC line-derived neurotrophic factor (GDNF) and mediating cell response to  
 CC GDNF. The GDNFR polypeptides and agonists can be used for treating  
 CC disorders associated with decreased activity of the respective  
 CC polypeptides. They can be used for treating neurodegenerative diseases  
 CC such as amyotrophic lateral sclerosis, Parkinson's disease, pituitary  
 CC schizophrenia, insomnia, tardive dyskinesia, hypertension, kidney  
 CC adenomas, hyperprolactinemia, thyroid tumour, renal disorders, kidney  
 CC failure, gut dysfunction, or for regeneration of cardiomyocytes,  
 CC epithelium or hepatocytes. Antagonists of the polypeptides can be used  
 CC for treating disorders associated with increased activity of the  
 CC respective polypeptides. The products can also be used for detection,  
 CC diagnosis and drug screening.  
 SQ Sequence 400 AA;

Query Match 88.9%; Score 40; DB 1; Length 400;  
 Best Local Similarity 30.8%; Pred. No. 3.06e+02;  
 Matches 4; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

Db 4 PLNRPRLPPVILM 16  
 | :|||  
 QY 2 PXXXXXXAVILM 14

RESULT 9  
 ID W65117 standard; Protein; 400 AA.  
 AC W65117;  
 DT 28-SEP-1998 (first entry)  
 DE Human GDNF alpha-3 receptor protein #2.  
 KW Glial cell line-derived neurotrophic factor alpha-3 receptor; GDNF;  
 KW treatment; neurodegenerative disease; Parkinson's disease; SMA;  
 KW amyotrophic lateral sclerosis; spinal muscular atrophy; nerve; trauma;  
 KW Huntington's disease; Alzheimer's disease; diabetic neuropathy; muscle;  
 KW muscular dystrophy; diagnostic.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT Protein 1..400  
 FT /label= GDNF alpha-3  
 FT /note= "partial sequence"  
 FT EP-846764-A2.  
 PN 10-JUN-1998.  
 PD 20-NOV-1997; 309375.  
 PF 09-MAY-1997; GB-009463.  
 PR 27-NOV-1996; GB-024677.  
 PA (SMIK ) SMITHKLINE BEECHAM PLC.  
 PI Lawrence GMP;  
 DR WPI: 98-299980/27.  
 DR N-PSDB: V35365.  
 PT New factor alpha 3 receptor polypeptide and e.g. DNA and agonists -  
 PT used to treat neuro degenerative diseases, muscular diseases and  
 PT nerve and muscle trauma and in diagnostic assays  
 PS Claim 13; Fig 4; 22pp; English.  
 CC This sequence represents a novel glial cell line-derived neurotrophic  
 CC factor alpha-3 receptor (GDNF alpha-3). This protein can be used to  
 CC treat e.g. neurodegenerative diseases (such as Parkinson's disease,  
 CC amyotrophic lateral sclerosis (ALS), spinal muscular atrophy (SMA),  
 CC Huntington's disease, Alzheimer's disease, diabetic neuropathy),  
 CC muscular diseases (including the muscular dystrophies) and nerve and  
 CC muscle trauma and in diagnostic assays for such conditions.  
 SQ Sequence 400 AA;

Query Match 88.9%; Score 40; DB 1; Length 400;  
 Best Local Similarity 30.8%; Pred. No. 3.06e+02;  
 Matches 4; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

Db 4 PLNRPRLPPVILM 16  
 | :|||  
 QY 2 PXXXXXXAVILM 14

RESULT 10  
 ID W65116 standard; Protein; 400 AA.  
 AC W65116;

PS Claim 11; Page 20; 31pp; Japanese.  
 CC This is the amino acid sequence of the human stromal cell protein  
 CC NT3-beta. The corresponding gene was isolated from cDNA library  
 CC generated from polyA+ purified human glial stem cell line T98G. The  
 CC probe was a fragment of the mouse NT3 gene (T63201). 80 positive clones  
 CC were isolated. Restriction digestion and subsequent separation by gel  
 CC electrophoresis revealed 3 major insert sizes: 1.9, 2.1 and 2.7 kb.  
 CC Sequence analysis showed 3 new NT3 gene sequences designated NT3-alpha,  
 CC (T63202), NT3-beta and NT3-gamma (T63204). The proteins can be used to  
 CC treat abnormal bone metabolism, irregular nervous system functioning or  
 CC insufficient growth of blood forming cells.  
 SQ Sequence 254 AA;

Query Match 88.9%; Score 40; DB 1; Length 254;  
 Best Local Similarity 38.5%; Pred. No. 3.06e+02;  
 Matches 5; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

Db 20 PCCLWLLGAVILM 32  
 |  
 |  
 QY 2 PXXXXXXAVILM 14

RESULT 5  
 ID W12373 standard; Protein; 362 AA.

AC W12373;  
 DT 08-MAY-1997 (first entry)  
 DE Human stromal cell NT3-alpha protein.  
 KW Mouse; stromal cell; NT3; human; abnormal bone metabolism; growth; blood;  
 KW nervous system functioning; glial stem cell.  
 OS Homo sapiens.

FH Key Location/Qualifiers  
 FT peptide 1..36 /label= "signal peptide"  
 FT protein 37..362 /note= "mature protein"

FN J08301898-A.  
 PD 19-NOV-1996.  
 PF 28-APR-1995; 128881.  
 PR 28-APR-1995; JP-128881.  
 PA (ONOV ) ONO PHARM CO LTD.  
 DR WPI: 97-048321/05.  
 DR N-PSDB; T63202.  
 PT DNA encoding polypeptide used to treat insufficient growth of  
 PT blood-forming cells - and abnormal bone metabolism  
 PS Claim 11; Page 16-17; 31pp; Japanese.  
 CC This is the amino acid sequence of the human stromal cell protein  
 CC NT3-alpha. The corresponding gene was isolated from cDNA library  
 CC generated from polyA+ purified human glial stem cell line T98G. The probe  
 CC was a fragment of the mouse NT3 gene (T63201). 80 positive clones were  
 CC isolated. Restriction digestion and subsequent separation by gel  
 CC electrophoresis revealed 3 major insert sizes: 1.9, 2.1 and 2.7 kb.  
 CC Sequence analysis showed 3 new NT3 gene sequences designated NT3-alpha,  
 CC NT3-beta (T63203) and NT3-gamma (T63204). The proteins can be used to  
 CC treat abnormal bone metabolism, irregular nervous system functioning or  
 CC insufficient growth of blood forming cells.  
 SQ Sequence 362 AA;

Query Match 88.9%; Score 40; DB 1; Length 362;  
 Best Local Similarity 38.5%; Pred. No. 3.06e+02;  
 Matches 5; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

Db 20 PCCLWLLGAVILM 32  
 |  
 |  
 QY 2 PXXXXXXAVILM 14

RESULT 6  
 ID W84185 standard; Protein; 378 AA.

AC W84185;  
 DT 25-MAR-1999 (first entry)  
 DE Glial cell line-derived neurotrophic factor receptor gamma 1.  
 KW Glial cell line-derived neurotrophic factor receptor gamma 1;  
 KW GDNFR-alpha; glial cell line-derived neurotrophic factor; GDNF;

KW neurodegenerative disease; amyotrophic lateral sclerosis; GDNFR-gamma;  
 KW Parkinson's disease; schizophrenia; insomnia; tardive dyskinesia;  
 KW hypertension; pituitary adenomas; hyperprolactinemia; thyroid tumour;  
 KW renal disorder; kidney failure; gut dysfunction; regeneration;  
 KW cardiomyocyte; epithelium; hepatocyte.

OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT Peptide 1..31 /label= signal\_peptide  
 FT Protein 32..378 /label= mature\_protein  
 FT Domain 32..360 /note= "extracellular domain"  
 FT Domain 361..378 /note= "transmembrane domain"

FN W09853069-A2.  
 PD 26-NOV-1998.

PF 20-MAY-1998; U10328.  
 PR 27-JUN-1997; US-884638.  
 PR 20-MAY-1997; US-047092.  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 PI Gentz RL, Hsu T, Ni J, Ruben SM, Young P;  
 DR WPI: 99-070150/06.  
 DR N-PSDB; V99333.

PT New isolated glial cell derived neurotrophic factor receptors - used  
 PT to develop products for treating e.g. neurodegenerative disorders,  
 PT schizophrenia, hypertension, tumours, renal disorders, kidney  
 PT failure or gut dysfunction  
 PS Claim 26; Fig 4A-C; 156pp; English.  
 CC The present sequence represents a glial cell line-derived neurotrophic  
 CC factor receptor gamma 1 (GDNFR-gamma). GDNFR-gamma shares high homology  
 CC with GDNFR-alpha, which is capable of complexing with glial cell  
 CC line-derived neurotrophic factor (GDNF) and mediating cell response to  
 CC GDNF. The GDNFR polypeptides and agonists can be used for treating  
 CC disorders associated with decreased activity of the respective  
 CC polypeptides. They can be used for treating neurodegenerative diseases  
 CC such as amyotrophic lateral sclerosis, Parkinson's disease,  
 CC schizophrenia, insomnia, tardive dyskinesia, hypertension, pituitary  
 CC adenomas, hyperprolactinemia, thyroid tumour, renal disorders, kidney  
 CC failure, gut dysfunction, or for regeneration of cardiomyocytes,  
 CC epithelium or hepatocytes. Antagonists of the polypeptides can be used  
 CC for treating disorders associated with increased activity of the  
 CC respective polypeptides. The products can also be used for detection,  
 CC diagnosis and drug screening.  
 SQ Sequence 378 AA;

Query Match 88.9%; Score 40; DB 1; Length 378;  
 Best Local Similarity 30.8%; Pred. No. 3.06e+02;  
 Matches 4; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

Db 4 PLNRPRLPPVILM 16  
 |  
 |  
 QY 2 PXXXXXXAVILM 14

RESULT 7  
 ID W84180 standard; Protein; 400 AA.

AC W84180;  
 DT 25-MAR-1999 (first entry)  
 DE A GDNFR-alpha-related protein 3 (GRR3).  
 KW Human; glial cell-line derived neurotrophic factor receptor;  
 KW GDNFR; glial cell line-derived neurotrophic factor; GDNF;  
 KW neurturin; signal transduction; dopaminergic nerve cell;  
 KW Parkinson's disease; Alzheimer's disease; amyotrophic lateral sclerosis;  
 KW neurological disorder; diabetes; glaucoma; sensory neuron;  
 KW retinal ganglion cell degeneration; sensory neuropathy;  
 KW retinopathy; gene therapy; GDNFR-related protein 3; GRR3.  
 OS Homo sapiens.  
 PN W09854213-A2.  
 PD 03-DEC-1998.  
 PF 27-APR-1998; U08485.  
 PR 30-MAY-1997; US-866354.  
 PA (AMGE-) AMGEN INC.

Db 186 PNASFVVYGVILM 198

\*\*\*\*\*  
M P S R C H  
\*\*\*\*\*  
(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Sat Apr 15 01:26:09 2000; MasPar time 3.21 Seconds  
Tabular output not generated. 103.395 Million cell updates/sec

Title: >US-08-452-843-23  
Description: (1-14) from US08452843.pep  
Perfect Score: 45  
Sequence: 1 XPXXXXXXAVILM 14

Scoring table: PAM 150  
Gap 15

Searched: 188963 seqs, 23686106 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: a-geneseqp  
1:geneseqp

Statistics: Mean 13.970; Variance 40.225; scale 0.347

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES					Pred. No.		
Result No.	Score	Query Match	Length	ID	Description		
1	45	100.0	485	1 W15280	AUX1 polypeptide invol	8.21e+01	
2	42	93.3	223	1 W87702	H. pylori GHPO 704 pro	1.82e+02	
3	40	88.9	44	1 R27731	SalF20R.	3.06e+02	
4	40	88.9	254	1 W12374	Human stromal cell NT3	3.06e+02	
5	40	88.9	362	1 W12373	Human stromal cell NT3	3.06e+02	
6	40	88.9	378	1 W84185	Glial cell line-derive	3.06e+02	
7	40	88.9	400	1 W84180	A GDNFR-alpha-related	3.06e+02	
8	40	88.9	400	1 W84186	Glial cell line-derive	3.06e+02	
9	40	88.9	400	1 W85117	Human GDNF alpha-3 rec	3.06e+02	
10	40	88.9	400	1 W65116	Human GDNF alpha-3 rec	3.06e+02	
11	40	88.9	400	1 W37463	Human Ret ligand RetL3	3.06e+02	
12	40	88.9	416	1 W70514	Candida glabrata IPC s	3.06e+02	
13	40	88.9	416	1 R27737	Sequence transcribed f	3.06e+02	
14	40	88.9	450	1 W12375	Human stromal cell NT3	3.06e+02	
15	40	88.9	724	1 W36797	Novel human gene, desi	3.06e+02	
16	40	88.9	818	1 W13386	Human protein ubiquiti	3.06e+02	
17	40	88.9	975	1 W93167	Human ZGBP1 protein.	3.06e+02	
18	39	86.7	36	1 R82536	IA beta chain fragment	3.96e+02	
19	39	86.7	43	1 R82539	IA beta chain fragment	3.96e+02	
20	39	86.7	59	1 W29209	I-Ag beta-chain insert	3.96e+02	
21	39	86.7	64	1 R82537	Hybrid IA beta chain f	3.96e+02	
22	39	86.7	67	1 R82540	Hybrid IA beta chain f	3.96e+02	
23	39	86.7	255	1 R82533	Hybrid IA beta chain.	3.96e+02	

24	39	86.7	298	1 R82538	Hybrid IA beta chain.	3.96e+02
25	39	86.7	354	1 R25853	MSH-dependent protein	3.96e+02
26	39	86.7	458	1 R98907	Vector SCE1-derived s1	3.96e+02
27	39	86.7	458	1 W29214	SCE1 single chain gene	3.96e+02
28	39	86.7	459	1 R98905	Vector SSC1-derived s1	3.96e+02
29	39	86.7	459	1 W29212	SSC1 single chain gene	3.96e+02
30	39	86.7	500	1 W29213	Vector SCT1-derived s1	3.96e+02
31	39	86.7	500	1 R98906	A NADPH oxidase derive	3.96e+02
32	39	86.7	943	1 W43039	HCMV Toledo strain UL1	5.12e+02
33	38	84.4	176	1 W05513	Human secreted protein	5.12e+02
34	38	84.4	245	1 W64220	Synechocystis sp. beta	5.12e+02
35	38	84.4	312	1 W99095	TriC enzyme.	5.12e+02
36	38	84.4	474	1 P70509	Fragment HGJ1789 of a	5.12e+02
37	38	84.4	2963	1 W56444	C. acetobutylicum inco	6.59e+02
38	37	82.2	318	1 W5502	CRFB4 protein.	6.59e+02
39	37	82.2	325	1 W52296	Human LAPH-1 protein s	6.59e+02
40	37	82.2	394	1 Y00876	Phosphoglycerate kinas	6.59e+02
41	37	82.2	417	1 R22095	PGK.	6.59e+02
42	37	82.2	418	1 R49247	CitH protein.	6.59e+02
43	37	82.2	426	1 W57833	B.diminuta pimelyl CoA	6.59e+02
44	37	82.2	525	1 W99453	Human butyrophilin.	6.59e+02
45	37	82.2	526	1 W97814		

ALIGNMENTS

RESULT 1  
ID W15280 standard; Protein; 485 AA.  
AC W15280; 1997 (first entry)  
DT 06-AUG-1997  
DE AUX1 polypeptide involved in auxin-related signalling pathway.  
KW AUX1 gene; auxin; signal transduction; transgenic plant; ripening;  
KW root gravitropism.  
OS Arabidopsis thaliana.  
PH Key Location/Qualifiers  
FT domain 51..69  
FT /label= Tm1  
FT /note= "transmembrane domain 1"  
FT domain 74..92  
FT /label= Tm2  
FT /note= "transmembrane domain 2"  
FT domain 141..157  
FT /label= Tm3  
FT /note= "transmembrane domain 3"  
FT domain 173..193  
FT /label= Tm4  
FT /note= "transmembrane domain 4"  
FT domain 199..213  
FT /label= Tm5  
FT /note= "transmembrane domain 5"  
FT domain 233..249  
FT /label= Tm6  
FT /note= "transmembrane domain 6"  
FT domain 269..288  
FT /label= Tm7  
FT /note= "transmembrane domain 7"  
FT domain 314..333  
FT /label= Tm8  
FT /note= "transmembrane domain 8"  
FT domain 358..376  
FT /label= Tm9  
FT /note= "transmembrane domain 9"  
FT domain 381..398  
FT /label= Tm10  
FT /note= "transmembrane domain 10"  
FT domain 430..444  
FT /label= Tm11  
FT /note= "transmembrane domain 11"  
FT peptide 203..210  
FT /label= Claim 9  
FT peptide 231..252  
FT /label= Claim 9  
FT peptide 251..260



#journal Nature (1997) 390:249-256  
 #title The complete genome sequence of the Gram-positive bacterium  
 Bacillus subtilis.  
 #cross-references MUID:98044033  
 #accession A69808  
 #status Preliminary; nucleic acid sequence not shown;  
 translation not shown  
 ##molecule\_type DNA  
 ##residues 1-351 ##label KUN  
 ##cross-references GB:299108; GB:AL009126; NID:g2633055; PID:e1182782;  
 PID:g2633116  
 ##experimental\_source strain 168

GENETICS  
 #gene yfkE  
 CLASSIFICATION #superfamily Ca2+/H+-exchanging protein  
 SUMMARY #length 351 #molecular-weight 37521 #checksum 4364  
 Query Match 88.9%; Score 40; DB 2; Length 351;  
 Best Local Similarity 38.5%; Pred. NO. 4.99e+01;  
 Matches 5; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

Db 308 PELVAMVSAVLIM 320  
 | | | | |  
 QY 2 PXXXXXXAVILM 14

RESULT 15  
 ENTRY A27595 #type fragment  
 TITLE H-2 class II histocompatibility antigen A-d beta chain  
 precursor - mouse (fragment)  
 ORGANISM #formal\_name Mus musculus #common\_name house mouse  
 DATE 08-Mar-1989 #sequence\_revision 08-Mar-1989 #text\_change  
 03-Jun-1994  
 ACCESSIONS A27595  
 REFERENCE A27595  
 #authors Griffith, I.J.; Nabavi, N.; Ghogawala, Z.; Chase, C.G.;  
 Rodriguez, M.; McKean, D.J.; Glimcher, L.H.  
 #journal J. Exp. Med. (1986) 167:541-555  
 #title Structural mutation affecting intracellular transport and  
 cell surface expression of murine class II molecules.  
 #cross-references MUID:88154743  
 #accession A27595

##molecule\_type mRNA  
 ##residues 1-27 ##label GRI  
 CLASSIFICATION #superfamily class II histocompatibility antigen;  
 immunoglobulin homology  
 KEYWORDS heterodimer  
 SUMMARY #length 27 #checksum 9401

Query Match 86.7%; Score 39; DB 2; Length 27;  
 Best Local Similarity 30.8%; Pred. NO. 8.11e+01;  
 Matches 4; Conservative 1; Mismatches 8; Indels 0; Gaps 0;  
 Db 6 PSLLLSAAVVVILM 18  
 | | | | |  
 QY 2 PXXXXXXAVILM 14

Search completed: Sat Apr 15 01:27:22 2000  
 Job time : 19 secs.

```
##residues 1-103 ##label SMI
##cross-references DDBJ:D11079; NID:g222717; PID:g1002303; PID:g222742
##experimental_source strain WR
REFERENCE A38550
#authors Howard, S.T.; Chan, Y.S.; Smith, G.L.
#journal Virology (1991) 180:633-647
#title Vaccinia virus homologues of the Shope fibroma virus inverted
terminal repeat proteins and a discontinuous ORF related to
the tumor necrosis factor receptor family.
#cross-references MIM:9111182
#accession B38550
##status preliminary
##molecule_type DNA
##residues 1-103 ##label HOW
##cross-references GB:M58052
##experimental_source strain WR
SUMMARY #length 103 #molecular-weight 12001 #checksum 8416
Query Match 88.9%; Score 40; DB 2; Length 103;
Best Local Similarity 30.8%; Pred. No. 4.99e+01;
Matches 4; Conservative 2; Mismatches 7; Indels 0; Gaps 0;
Db 76 PPHYTVIPVILM 88
QY 2 PXXXXXXXAVILM 14
RESULT 12
ENTRY F70943 #type complete
TITLE hypothetical protein RV2039c - Mycobacterium tuberculosis
(strain H37RV)
ORGANISM #formal_name Mycobacterium tuberculosis
DATE 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change
17-Jul-1998
ACCESSION F70943
REFERENCE A70500
#authors Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher,
C.; Harris, D.; Gordon, S.V.; Eiglmeyer, K.; Gas, S.; Barry
III, C.E.; Tekala, F.; Badcock, K.; Basham, D.; Brown, D.;
Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.;
Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;
Hornsby, T.; Jagels, K.; Krogh, A.; McLean, J.; Moule, S.;
Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.;
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.;
Skellton, S.; Squares, S.; Squires, R.; Sulston, J.E.;
Taylor, K.; Whitehead, S.; Barrell, B.G.
#journal Nature (1998) 393:537-544
#title Deciphering the biology of Mycobacterium tuberculosis from
the complete genome sequence.
#cross-references MIM:9829598
#accession F70943
##status preliminary; nucleic acid sequence not shown;
translation not shown
##molecule_type DNA
##residues 1-280 ##label COL
##cross-references GB:AL021899; GB:AL123456; NID:g3242282; PID:e1252018;
PID:g2896776
##experimental_source strain H37RV
GENETICS
#gene RV2039c
SUMMARY #length 280 #molecular-weight 31607 #checksum 4091
Query Match 88.9%; Score 40; DB 2; Length 280;
Best Local Similarity 38.5%; Pred. No. 4.99e+01;
Matches 5; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
Db 122 PLFVTVIPVILM 134
QY 2 PXXXXXXXAVILM 14
RESULT 13
ENTRY T01949 #type complete
```

```
TITLE hypothetical protein F1104.14 - Arabidopsis thaliana
ORGANISM #formal_name Arabidopsis thaliana #common_name mouse-ear
cress
DATE 26-Feb-1999 #sequence_revision 26-Feb-1999 #text_change
24-Mar-1999
ACCESSION T01949
REFERENCE Z14466
#authors Abu-Threideh, J.; Stoneking, T.; Langston, Y.; Trevaskis, E.
#journal submitted to the EMBL Data Library, October 1998
#description The sequence of A. thaliana F1104.
#accession T01949
##status translated from GB/EMBL/DDBJ
##molecule_type DNA
##residues 1-340 ##label ABU
##cross-references EMBL:AF096370; NID:g3695372; PID:g3695385
GENETICS
#map_position 4
#introns 62/2; 101/2; 230/3; 281/2
#note F1104.14
SUMMARY #length 340 #molecular-weight 37483 #checksum 3937
Query Match 88.9%; Score 40; DB 2; Length 340;
Best Local Similarity 30.8%; Pred. No. 4.99e+01;
Matches 4; Conservative 2; Mismatches 7; Indels 0; Gaps 0;
Db 251 PVVSSFSFVPLM 263
QY 2 PXXXXXXXAVILM 14
RESULT 14
ENTRY A69808 #type complete
TITLE H+/Ca2+ exchanger homolog yfke - Bacillus subtilis
ORGANISM #formal_name Bacillus subtilis
DATE 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change
24-Sep-1998
ACCESSION A69808
REFERENCE A69590
#authors Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.;
Alloni, G.; Azevedo, V.; Bertero, M.G.; Bessieres, P.;
Bolotin, A.; Borchert, S.; Boriss, R.; Boursier, L.; Brans,
A.; Braun, M.; Brignell, S.C.; Bron, S.; Brouillet, L.; Bruns,
Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
Choi, S.K.; Codani, J.J.; Connerton, I.F.; Cummings, N.J.;
Daniel, R.A.; Denizot, F.; Devine, K.M.; Duesterhoeft, A.;
Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.;
Fabret, C.; Ferrari, E.; Foulger, D.; Fritz, C.; Fujita,
M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galleron, N.; Ghim,
S.Y.; Glaser, P.; Goffeau, A.; Golightly, E.J.; Grandi, G.;
Guiseppi, G.; Guy, B.J.; Haga, K.; Haiech, J.; Harwood,
C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.;
Hullo, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karamata, D.;
Kasahara, Y.; Klaerr-Blanchard, M.; Klein, C.; Kobayashi,
Y.; Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.;
Kurita, K.; Lapidus, A.; Lardinois, S.; Lauber, J.;
Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.;
Maueel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno,
M.; Moestl, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly,
V.; Pohl, T.M.; Portetelle, D.; Porwollik, S.; Prescott,
A.M.; Presecan, E.; Pujic, P.; Purnelle, B.; Rapoport, G.;
Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.;
Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon, E.;
Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.;
Sekowska, A.; Seror, S.J.; Serror, P.; Shin, B.S.; Soldo,
B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Takanashi, H.;
Takemaru, K.; Takeuchi, M.; Tamakoshi, A.; Tanaka, T.;
Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.;
Vandenbol, M.; Vannier, F.; Vassarotti, A.; Viari, A.;
Wambutt, R.; Wedler, E.; Wedler, H.; Weitzenecker, T.;
Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto,
K.; Yata, K.; Yoshida, K.; Yoshikawa, H.F.; Zumbstein, E.;
Yoshikawa, H.; Danchin, A.
```

```
##molecule_type DNA
##residues 1-312 ##label KLE
##cross-references GB:AE001008; GB:AE000782; NID:g2689331; PID:g2649184;
TIGR:AF1392
SUMMARY #length 312 #molecular-weight 33392 #checksum 5309

Query Match 91.1%; Score 41; DB 2; Length 312;
Best Local Similarity 38.5%; Pred. No. 3.04e+01;
Matches 5; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Db 92 PLIRGASVILM 104
|
|
|
Qy 2 PXXXXXXAVILM 14

RESULT 8
ENTRY D71416 #type complete
TITLE Probable PDR5-like ABC transporter - Arabidopsis thaliana
ORGANISM #formal_name Arabidopsis thaliana #common_name mouse-ear
cress
#variety columbia
DATE 03-Aug-1998 #sequence_revision 03-Aug-1998 #text_change
19-Feb-1999
D71416
A71400
Bevan, M.; Bancroft, I.; Bent, E.; Love, K.; Goodman, H.;
Dean, C.; Bergkamp, R.; Dirkse, W.; Van Staveren, M.;
Stiekema, W.; Drost, L.; Ridley, P.; Hudson, S.A.; Patel,
K.; Murphy, G.; Piffanelli, P.; Wedler, H.; Wedler, E.;
Wambutt, R.; Weitzenegger, T.; Pohl, T.M.; Terryn, N.;
Gielen, J.; Villarroel, R.; De Clerck, R.; Van Montagu, M.;
Lecharny, A.; Auborg, S.; Gy, I.; Kreis, M.; Lao, N.;
Kavanagh, T.; Hempel, S.; Kotter, P.; Entian, K.D.; Rieger,
M.; Schaeffer, M.; Funk, B.; Mueller-Auer, S.; Silvey, M.;
James, R.; Montfort, A.; Pons, A.; Puigdomenech, P.; Douka,
A.; Voukelatou, E.; Milioni, D.; Hatzopoulos, P.;
Piravandi, E.; Obermaier, B.; Hilbert, H.; Duesterhoft, A.;
Moore, T.; Jones, J.D.G.; Eneva, T.; Palme, K.; Benes, V.;
Rechman, S.; Ansoorge, W.; Cooke, R.; Berger, C.; Deisen,
M.; Voet, M.; Volckaert, G.; Mewes, H.W.; Klosterman, S.;
Schueller, C.; Chalwatzis, N.
#journal Nature (1998) 391:485-488
#title Analysis of 1.9 Mb of contiguous sequence from chromosome 4
of Arabidopsis thaliana.
#cross-references MUID:98121113
#accession D71416
#status preliminary; nucleic acid sequence not shown;
translation not shown

##molecule_type DNA
##residues 1-1177 ##label BEV
##cross-references GB:297338; NID:g2244870; PID:g326915; PID:g2244881
GENETICS
#map_position 4COP9-4G3845
CLASSIFICATION #superfamily ATP-binding cassette homology
FEATURE
439-653 #domain ATP-binding cassette homology #label ABC2
SUMMARY #length 1177 #molecular-weight 134263 #checksum 2760

Query Match 91.1%; Score 41; DB 2; Length 1177;
Best Local Similarity 38.5%; Pred. No. 3.04e+01;
Matches 5; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Db 106 PETFELEDDVILM 118
|
|
|
Qy 2 PXXXXXXAVILM 14

RESULT 9
ENTRY T02644 #type complete
TITLE Probable ABC transport protein - Arabidopsis thaliana
ORGANISM #formal_name Arabidopsis thaliana #common_name mouse-ear
cress
DATE 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change
```

```
24-Mar-1999
T02644
Z14685
Rounsley, S.D.; Ronning, C.M.; Lin, X.; Ketchum, K.A.;
Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; Kaul, S.; Mason,
T.M.; Kerlavage, A.R.; Adams, M.D.; Somerville, C.R.;
Venter, J.C.
#submission submitted to the EMBL Data Library, August 1998
#description Arabidopsis thaliana chromosome II BAC F12C20 genomic
sequence.
#accession T02644
#status preliminary; translated from GB/EMBL/DBJ
#molecule_type DNA
##residues 1-1420 ##label ROU
##cross-references EMBL:AC005168; NID:g3426033; PID:g3426037
GENETICS
#map_position 2
#introns 92/2; 132/3; 162/2; 190/3; 244/1; 269/3; 287/3; 318/1; 418/3;
512/3; 618/2; 672/1; 706/3; 760/1; 801/1; 831/3; 942/3;
1039/3; 1067/3; 1112/2; 1188/2; 1245/3; 1330/3
F12C20.5
#note #length 1420 #molecular-weight 161263 #checksum 890
SUMMARY

Query Match 91.1%; Score 41; DB 2; Length 1420;
Best Local Similarity 38.5%; Pred. No. 3.04e+01;
Matches 5; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Db 369 PETFELEDDVILM 381
|
|
|
Qy 2 PXXXXXXAVILM 14

RESULT 10
ENTRY A42523 #type complete
TITLE A53R protein - vaccinia virus (strain Copenhagen)
ORGANISM #formal_name vaccinia virus
#note host Homo sapiens (man)
DATE 09-Nov-1990 #sequence_revision 09-Nov-1990 #text_change
08-Apr-1994
A42523
A33172
Johnson, G.P.
#authors submitted to GenBank, June 1990
#submission #accession A42523
#status preliminary
#molecule_type DNA
#residues 1-103 ##label JOH
SUMMARY #length 103 #molecular-weight 12032 #checksum 8457

Query Match 88.9%; Score 40; DB 2; Length 103;
Best Local Similarity 30.8%; Pred. No. 4.99e+01;
Matches 4; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

Db 76 PPHYTTIPILM 88
|
|
|
Qy 2 PXXXXXXAVILM 14

RESULT 11
ENTRY JQ1791 #type complete
TITLE Salp16R protein - vaccinia virus
ALTERNATE_NAMES Salp16R 12k protein
ORGANISM #formal_name vaccinia virus
DATE 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change
09-Sep-1997
JQ1791; B38550
#accessions JQ1791; B38550
REFERENCE JQ1767
#authors Smith, G.L.; Chan, Y.S.; Howard, S.T.
#journal J. Gen. Virol. (1991) 72:1349-1376
#title Nucleotide sequence of 42kbp of vaccinia virus strain WR from
near the right inverted terminal repeat.
#accession JQ1791
#molecule_type DNA
```

Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.; Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.  
 Nature (1997) 390:249-256  
 The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.  
 accession C69787  
 #cross-references EMBL:98044033  
 #status preliminary; nucleic acid sequence not shown; translation not shown  
 #molecule\_type DNA  
 #residues 1-254 #label KUN  
 #cross-references GB:299107; GB:AL009126; NID:g2632866; PID:e1182578; PID:g2632912  
 #experimental\_source strain 168

GENETICS  
 #gene ydlJ  
 CLASSIFICATION #superfamily conserved hypothetical protein HI0188  
 SUMMARY #length 254 #molecular-weight 29056 #checksum 4722

Query Match 93.3%; Score 42; DB 2; Length 254;  
 Best Local Similarity 38.5%; Pred. No. 1.84e+01;  
 Matches 5; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

Db 164 PFGLFQMPVILM 176  
 |  
 :|||  
 QY 2 PXXXXXXAVILM 14

RESULT 6  
 ENTRY  
 TITLE SCT1 protein - yeast (Saccharomyces cerevisiae)  
 ALTERNATE\_NAMES protein YBL011w; protein YBL0309; protein YBL0315  
 ORGANISM #formal\_name Saccharomyces cerevisiae  
 DATE 23-Apr-1993 #sequence\_revision 23-Apr-1993 #text\_change 17-Mar-1999

ACCESSIONS S25330; S41216; S45745; S45744; JC4182  
 REFERENCE S25330  
 #authors Skala, J.; van Dyck, L.; Purnelle, B.; Goffeau, A.  
 #journal Yeast (1992) 8:777-785  
 #title The sequence of an 8 kb segment on the left arm of chromosome II from *Saccharomyces cerevisiae* identifies five new open reading frames of unknown functions, two tRNA genes and two transposable elements.  
 #cross-references MUID:93070615  
 #accession S25330  
 #molecule\_type DNA  
 #residues 1-759 #label SKA  
 #cross-references EMBL:S47701  
 #experimental\_source strain S288C  
 #note the authors did not translate the codon GAA for residue 712

REFERENCE S25326  
 #authors Delaveau, T.; Jacq, C.; Perea, J.  
 #journal Yeast (1992) 8:761-768  
 #title Sequence of a 12.7 kb segment of yeast chromosome II identifies a PDR-like gene and several new open reading frames.  
 #cross-references MUID:93070613  
 #accession S41216  
 #status translation not shown  
 #molecule\_type DNA  
 #residues 609-759 #label DEL  
 #cross-references EMBL:S47695; NID:g259049; PID:g1680404  
 #experimental\_source strain S288C

REFERENCE S45745  
 #authors Goffeau, A.; Jonniaux, J.L.; Purnelle, B.; Skala, J.; de Wergifosse, P.; van Dyck, L.  
 #submission submitted to the Protein Sequence Database, August 1994  
 #accession S45745  
 #molecule\_type DNA  
 #residues 1-610 #label GOF  
 #cross-references EMBL:Z35773; MIPS:YBL011w

REFERENCE S45736  
 #authors Delaveau, T.; Jacq, C.; Perea, J.  
 #submission submitted to the Protein Sequence Database, August 1994  
 #accession S45744  
 #molecule\_type DNA  
 #residues 578-759 #label DE2  
 #cross-references EMBL:Z35773; MIPS:YBL011w  
 #accession JC4182

REFERENCE JC4182  
 #authors Matsushita, M.; Nikawa, J.I.  
 #journal J. Biochem. (1995) 117:447-451  
 #title Isolation and characterization of a SCT1 gene which can suppress a choline-transport mutant of *Saccharomyces cerevisiae*.  
 #accession JC4182  
 #molecule\_type DNA  
 #residues 1-9,'S',11-87,'R',89-124,'A',126-729,'S',731-759 #label MAT

GENETICS  
 #gene SCT1  
 #map\_position 2L  
 YBL011w  
 choline transport; glycoprotein; phosphoprotein; transmembrane protein

KEYWORDS  
 FEATURE 113,468,619  
 441-457  
 494-510  
 532-548  
 686-711  
 21

#domain transmembrane #status predicted #label TM1\  
 #domain transmembrane #status predicted #label TM2\  
 #domain transmembrane #status predicted #label TM3\  
 #region PST sequence\  
 #binding\_site carbohydrate (Asn) (covalent) #status predicted\  
 #binding\_site phosphate (Ser) (covalent) (by protein kinase A) #status predicted

SUMMARY #length 759 #molecular-weight 85694 #checksum 9658

Query Match 93.3%; Score 42; DB 2; Length 759;  
 Best Local Similarity 38.5%; Pred. No. 1.84e+01;  
 Matches 5; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

Db 89 PHANQFVDPVILM 101  
 |  
 :|||  
 QY 2 PXXXXXXAVILM 14

RESULT 7  
 ENTRY  
 TITLE branched-chain amino acid ABC transporter, permease protein (brad-4) homolog - *Archaeoglobus fulgidus*  
 ORGANISM #formal\_name *Archaeoglobus fulgidus*  
 DATE 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 05-Jun-1998

ACCESSIONS G69423  
 REFERENCE A69250  
 #authors Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson, R.J.; Gwinn, M.; Hickey, E.K.; Peterson, J.D.; Richardson, D.L.; Kervilave, A.R.; Graham, D.E.; Kyrpides, N.C.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.; Dougherty, B.A.; McKenny, K.; Adams, M.D.; Loftus, B.; Peterson, S.; Reich, C.I.; McNeil, L.K.; Badger, J.H.; Glodek, A.; Zhou, L.; Overbeck, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.; Utterback, T.; Cotton, M.D.; Spriggs, T.; Arriach, P.; Kaine, B.P.; Sykes, S.M.; Sadow, P.W.; D'Andrea, K.P.; Bowman, C.; Fujii, C.; Garland, S.A.; Mason, T.M.; Olsen, G.J.; Fraser, C.M.; Smith, H.O.; Woese, C.R.; Venter, J.C.  
 Nature (1997) 390:364-370  
 The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeon *Archaeoglobus fulgidus*.  
 #cross-references MUID:98049343  
 #accession G69423  
 #status preliminary; nucleic acid sequence not shown; translation not shown

```

Y.
#journal Science (1997) 277:1453-1462
#title The complete genome sequence of Escherichia coli K-12.
#cross-references MUID:97426617
#accession G55004
#status preliminary; nucleic acid sequence not shown;
translation not shown
#molecule_type DNA
#residues 1-331 #label BLAT
#cross-references GB:AE000321; GB:U00096; NID:g1788659; PID:g1788661;
UMGP:b2321
#experimental_source strain K-12, substrain MG1655
GENETICS
#gene div
SUMMARY
#length 331 #molecular-weight 36668 #checksum 690
Query Match 95.6%; Score 43; DB 2; Length 331;
Best Local Similarity 38.5%; Pred. No. 1.10e+01;
Matches 5; Conservative 1; Mismatches 7; Indels 0; Gaps 0;
Db 313 PALMILLVAILM 325
QY 2 PXXXXXXXAVILM 14
I:||||
I:||||

RESULT 3
ENTRY B54666 #type complete
TITLE glutamine ABC transporter, permease protein - Helicobacter
ORGANISM pylori (strain 26695)
#formal_name Helicobacter pylori
DATE 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change
10-Oct-1997
ACCESSIONS B54666
REFERENCE A64520
#authors Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.;
Sutton, G.G.; Fleischmann, R.D.; Ketchum, K.A.; Klenk,
H.P.; Gill, S.; Dougherty, B.A.; Nelson, K.; Quackenbush,
J.; Zhou, L.; Kirkness, E.F.; Peterson, S.; Loftus, B.;
Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.;
McKenney, K.; Fitzgerald, L.M.; Lee, N.; Adams, M.D.;
Hickey, E.K.; Berg, D.E.; Gocayne, J.D.; Utterback, T.R.;
Peterson, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.;
Fuji, C.; Bowman, C.; Watthey, L.; Wallin, E.; Hayes,
W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser,
C.M.; Venter, J.C.
#journal Nature (1997) 388:539-547
#title The complete genome sequence of the gastric pathogen
Helicobacter pylori.
#cross-references MUID:97394467
#accession B54666
#status preliminary; nucleic acid sequence not shown;
translation not shown
#molecule_type DNA
#residues 1-223 #label TOM
#cross-references GB:AE000623; GB:AE000511; NID:g2314327; PID:g2314329;
TIGR:HP1170
SUMMARY
#length 223 #molecular-weight 24751 #checksum 7411
Query Match 93.3%; Score 42; DB 2; Length 223;
Best Local Similarity 38.5%; Pred. No. 1.84e+01;
Matches 5; Conservative 1; Mismatches 7; Indels 0; Gaps 0;
Db 186 PNASFVYGVILM 198
QY 2 PXXXXXXXAVILM 14
I:||||
I:||||

RESULT 4
ENTRY E71849 #type complete
TITLE amino acid ABC transporter, permease protein - Helicobacter
ORGANISM pylori (strain J99)
#formal_name Helicobacter pylori
#variety strain J99

```

```

DATE 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change
12-Feb-1999
E71849
REFERENCE
#authors Alm, R.A.; Ling, L.S.L.; Molir, D.T.; King, B.L.; Brown, E.D.;
Doig, P.C.; Smith, D.R.; Noonan, B.; Guild, B.C.; deJonge,
B.L.; Carmel, G.; Tummino, P.J.; Caruso, A.;
Uria-Nickelsen, M.; Mills, D.M.; Ives, C.; Gibson, R.;
Marberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis,
G.F.; Trust, T.J.
#journal Nature (1999) 397:176-180
#title Genomic sequence comparison of two unrelated isolates of the
human gastric pathogen Helicobacter pylori.
#cross-references MUID:99120557
#accession E71849
#status preliminary
#molecule_type DNA
#residues 1-223 #label ABN
#cross-references GB:AE001537; GB:AE001439; NID:g4155687; PID:g4155695
#experimental_source strain J99
GENETICS
#gene jhp1097
SUMMARY
#length 223 #molecular-weight 24749 #checksum 8892
Query Match 93.3%; Score 42; DB 2; Length 223;
Best Local Similarity 38.5%; Pred. No. 1.84e+01;
Matches 5; Conservative 1; Mismatches 7; Indels 0; Gaps 0;
Db 186 PNASFVYGVILM 198
QY 2 PXXXXXXXAVILM 14
I:||||
I:||||

RESULT 5
ENTRY C69787 #type complete
TITLE conserved hypothetical protein ydiJ - Bacillus subtilis
ORGANISM #formal_name Bacillus subtilis
DATE 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change
24-Sep-1998
C69787
REFERENCE A69580
#authors Kunat, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.;
Alloni, G.; Azevedo, V.; Bertero, M.G.; Bessieres, P.;
Bolotin, A.; Borchert, S.; Boriss, R.; Boursier, L.; Brans,
A.; Braun, M.; Brignell, S.C.; Bron, S.; Brouillet, S.;
Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
Choi, S.K.; Codani, J.J.; Conneton, I.F.; Cummings, N.J.;
Daniel, R.A.; Denizot, F.; Devine, K.M.; Duesterhoeft, A.;
Ehrlich, S.D.; Emerson, P.T.; Entlan, K.D.; Errington, J.;
Fabret, C.; Ferrari, E.; Foulger, D.; Fritz, C.; Fujita,
M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galleron, N.; Ghim,
S.Y.; Glaser, P.; Goffeau, A.; Golightly, E.J.; Grandi, G.;
Guiseppi, G.; Guy, B.J.; Haga, K.; Haiech, J.; Harwood,
C.R.; Hentat, A.; Hilbert, H.; Holsappel, S.; Hosono, S.;
Hullo, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karamata, D.;
Kasahara, Y.; Klaerr-Blanchard, M.; Klein, C.; Kobayashi,
Y.; Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.;
Kurita, K.; Lapidus, A.; Lardinois, S.; Lauber, J.;
Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.;
Mauel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno,
M.; Moesti, D.; Nakai, S.; Noback, M.; Noone, B.S.; O'Reilly,
M.; Ogawa, K.; Ogihara, A.; Oudega, B.; Park, S.H.; Parro,
V.; Pohl, T.M.; Portetle, D.; Porwoll, S.; Prescott,
A.M.; Prescan, E.; Pujic, P.; Purnelle, B.; Rapoport, G.;
Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.;
Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon, E.;
Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.;
Sekowska, A.; Seror, S.J.; Serror, P.; Shin, B.S.; Soldo,
B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Takahashi, H.;
Takemaru, K.; Takeuchi, M.; Tamakoshi, A.; Tanaka, T.;
Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.;
Vandenbol, M.; Vannier, F.; Vassarotti, A.; Viari, A.;
Wambutt, R.; Wedler, E.; Wedler, H.; Weitzenecker, T.;

```

\*\*\*\*\*

MPARCH\_PP protein - protein database search, using Smith-Waterman algorithm

\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

Run on: Sat Apr 15 01:27:03 2000; MasPar time 3.31 Seconds

Tabular output not generated. 169.531 Million cell updates/sec

Title: >US-08-452-843-23  
Description: (1-14) from US08452843.pep  
Perfect Score: 45  
Sequence: 1 PXXXXXXXAVILM 14

Scoring table: PAM 150  
Gap 15

Searched: 122810 seqs, 40068593 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: p1r62  
1:p1r1 2:p1r2 3:p1r3 4:p1r4

Statistics: Mean 20.311; Variance 22.428; scale 0.906

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	45	100.0	485	2	hypothetical protein	3.84e+00
2	43	95.6	331	2	Div protein - Escheri	1.10e+01
3	42	93.3	223	2	glutamine ABC transpo	1.84e+01
4	42	93.3	223	2	amino acid ABC transp	1.84e+01
5	42	93.3	254	2	conserved hypothetical	1.84e+01
6	42	93.3	759	2	SC11 protein - yeast	1.84e+01
7	41	91.1	312	2	branched-chain amino	3.04e+01
8	41	91.1	1177	2	probable PDR5-like AB	3.04e+01
9	41	91.1	1420	2	probable ABC transpor	3.04e+01
10	40	88.9	103	2	A53R protein - vacci	4.99e+01
11	40	88.9	103	2	SalF16R protein - vac	4.99e+01
12	40	88.9	280	2	hypothetical protein	4.99e+01
13	40	88.9	340	2	Hypothetical protein	4.99e+01
14	40	88.9	351	2	H+/Ca2+ exchanger hom	4.99e+01
15	39	86.7	27	2	H-2 class II histocom	8.11e+01
16	39	86.7	112	2	hypothetical protein	8.11e+01
17	39	86.7	142	2	probable membrane pro	8.11e+01
18	39	86.7	237	2	stearoyl-CoA desatur	8.11e+01
19	39	86.7	263	1	H-2 class II histocom	8.11e+01
20	39	86.7	263	1	H-2 class II histocom	8.11e+01
21	39	86.7	263	1	H-2 class II histocom	8.11e+01
22	39	86.7	263	1	H-2 class II histocom	8.11e+01
23	39	86.7	263	1	class II histocompati	8.11e+01

24	39	86.7	255	1	HLMSAB	MHC class II histocom	8.11e+01
25	39	86.7	255	1	HLMSQB	MHC class II histocom	8.11e+01
26	39	86.7	265	2	I48656	histocompatibility cl	8.11e+01
27	39	86.7	355	2	A32115	stearoyl-CoA desatur	8.11e+01
28	39	86.7	358	2	A24699	stearoyl-CoA desatur	8.11e+01
29	39	86.7	358	2	A36507	stearoyl-CoA desatur	8.11e+01
30	39	86.7	567	2	H69145	sensory transporter	8.11e+01
31	39	86.7	573	2	E69802	ABC transporter (ATP-	8.11e+01
32	39	86.7	806	2	E69424	hypothetical protein	8.11e+01
33	39	86.7	944	2	T00265	respiratory burst oxi	8.11e+01
34	39	86.7	2238	1	RRVUB5	genome polyprotein -	8.11e+01
35	38	84.4	150	2	S74855	hypothetical protein	1.30e+02
36	38	84.4	447	2	G70030	amino acid permease h	1.30e+02
37	38	84.4	461	2	A30222	hypothetical D15K21 p	1.30e+02
38	38	84.4	474	2	E24723	trpC protein - Coryne	1.30e+02
39	38	84.4	499	2	S47160	DNA-directed RNA poly	1.30e+02
40	38	84.4	508	2	JC6200	cytochrome P450scs -	1.30e+02
41	38	84.4	520	2	C64213	ribose transport syst	1.30e+02
42	38	84.4	810	2	S67050	probable membrane pro	1.30e+02
43	38	84.4	1110	2	I59370	guanylate cyclase (EC	1.30e+02
44	38	84.4	1753	2	T00350	hypothetical protein	1.30e+02
45	38	84.4	3951	1	VFIHBI	F1 protein - avian in	1.30e+02

ALIGNMENTS

RESULT 1

ENTRY T01244  
TITLE #type complete  
ORGANISM #formal\_name Arabidopsis thaliana #common\_name mouse-ear

DATE 12-Feb-1999 #sequence\_revision 12-Feb-1999 #text\_change 12-Feb-1999

ACCESSIONS T01244  
REFERENCE T14284  
#authors Rounsley, S.D.; Kaul, S.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; Mason, T.M.; Kerlavage, A.R.; Adams, M.D.; Somerville, C.R.; Venter, J.C.

#submission submitted to the EMBL Data Library, July 1998  
#description Arabidopsis thaliana chromosome II BAC F16M14 genomic sequence.  
#accession T01244  
#status preliminary; translated from GB/EMBL/DBJ  
#molecule\_type DNA  
#residues 1-485 #label ROU  
#cross-references EMBL:AC003028; NID:g33335356; PID:g33335360

GENETICS  
#map\_position II  
#introns 59/3; 121/3; 159/2; 220/3; 253/2; 322/3; 408/3  
#note F16M14.5

SUMMARY  
#length 485 #molecular-weight 54059 #checksum 3498

Query Match 100.0%; Score 45; DB 2: Length 485;  
Best Local Similarity 46.2%; Pred. No. 3.84e+00;  
Matches 6; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Db 306 PKNAWRDAAVILM 318  
QY 2 PXXXXXXXAVILM 14

RESULT 2  
ENTRY G65004  
TITLE #type complete  
ORGANISM #formal\_name Escherichia coli (strain K-12)  
DATE 12-Sep-1997 #sequence\_revision 17-Sep-1997 #text\_change 14-Nov-1997

ACCESSIONS G65004  
REFERENCE A64720  
#authors Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.; Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.; Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao,

RA KASAI H., KASHIMOTO K., KIM S., KIMURA S., KITAGAWA M.,  
 RA KITAKAWA M., MAKINO K., MASUDA S., MIKI T., MIZOBUCHI K., MORI H.,  
 RA MOTOMURA K., NAKAMURA Y., NASHIMOTO H., NISHIO Y., OSHIMA T.,  
 RA SAITO N., SAMEI G., SEKI Y., TAGAMI H., TAKEMOTO K., WADA C.,  
 RA YAMAMOTO Y., YANO M.,  
 RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.  
 CC -!- CATALYTIC ACTIVITY: 5-O-(-1-CARBOXYVINYL)-3-PHOSPHOSHIKIMATE -  
 CC CHORISMATE + ORTHOPHOSPHATE.  
 CC -!- COFACTOR: REDUCED FLAVIN.  
 CC -!- PATHWAY: SEVENTH STEP IN THE BIOSYNTHESIS OF CHORISMATE WITHIN  
 CC THE BIOSYNTHESIS OF AROMATIC AMINO ACIDS (THE SHIKIMATE PATHWAY).  
 CC -!- SUBUNIT: HOMOTETRAMER.  
 CC -!- SIMILARITY: BELONGS TO THE CHORISMATE SYNTHASE FAMILY.  
 CC -!- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN FROM POSITION 334  
 CC ONWARD AND IS SHORTER (355 AA) DUE TO A FRAMESHIFT.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; M27714; AA23487.1; -  
 DR EMBL; Y00720; CAA68707.1; ALT\_FRAME.  
 DR EMBL; M33021; AA23488.1; ALT\_FRAME.  
 DR EMBL; AE000321; AAC75389.1; -  
 DR EMBL; D90863; CAB22102.1; -  
 DR EMBL; D90864; CAB22116.1; -  
 DR PIR; S00509; SYECKR.  
 DR ECOGENE; EGI0075; AROC.  
 DR PROSITE; PS00787; CHORISMATE\_SYNTHASE\_1; 1.  
 DR PROSITE; PS00788; CHORISMATE\_SYNTHASE\_2; 1.  
 DR PROSITE; PS00789; CHORISMATE\_SYNTHASE\_3; 1.  
 DR PFAM; PF01264; Chorismate\_synt; 1.  
 KW Lyase; Aromatic amino acid biosynthesis.  
 FT INIT\_MET 0  
 SQ SEQUENCE 360 AA; 39006 MW; 994660B7 CRC32;  
  
 Query Match 91.1%; Score 41; DB 1; Length 360;  
 Best Local Similarity 33.3%; Pred. No. 1.08e+01;  
 Matches 4; Conservative 2; Mismatches 6; Indels 0; Gaps 0;  
  
 Db 329 PIAEAMLAIVLM 340  
 |  
 |::||  
 QY 2 PXXXXXAVILM 13

Search completed: Sat Apr 15 01:21:06 2000  
 Job time : 40 secs.

CC -1- FUNCTION: REGULATES A SIGNAL TRANSDUCTION PATHWAY LINKING PLASMA  
CC MEMBRANE RECEPTORS TO THE ASSEMBLY OF FOCAL ADHESIONS AND ACTIN  
CC STRESS FIBERS.  
CC -1- SIMILARITY: BELONGS TO THE RHO FAMILY IN THE RAS SUPERFAMILY.  
CC  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; M12174; AAA36565.1; -  
CC EMBL; X06820; CAA29968.1; -  
CC EMBL; M74295; AAA42040.1; -  
CC EMBL; X99963; CAA58228.1; -  
CC PIR; A01372; TVHURH.  
CC PIR; A39727; TVTRRH.  
CC HSSP; P06749; 1A2B.  
CC MIM; 165370; -  
CC MGD; MGI:107949; ARHB.  
CC PFAM; PF00071; ras; 1.  
CC  
CC Proto-oncogene; GTP-binding; Prenylation; Lipoprotein.  
CC FT NP\_BIND 12 19 GTP (BY SIMILARITY).  
CC FT NP\_BIND 59 63 GTP (BY SIMILARITY).  
CC FT NP\_BIND 117 120 GTP (BY SIMILARITY).  
CC FT DOMAIN 34 42 EFFECTOR REGION (POTENTIAL).  
CC FT LIPID 193 193 GERANYL-GERANYL (BY SIMILARITY).  
CC SQ SEQUENCE 196 AA; 22123 MW; 7E371891 CRC32;  
  
Query Match 91.1%; Score 41; DB 1; Length 196;  
Best Local Similarity 41.7%; Pred. No. 1.08e+01;  
Matches 5; Conservative 0; Mismatches 7; Indels 0; Gaps 0;  
  
Db 71 PLSYPDTVDVILM 82  
QY 2 PXXXXXXAVILM 13  
  
RESULT 14  
ID KCDA\_DROME STANDARD; PRT; 354 AA.  
AC P16910;  
DT 01-AUG-1990 (Rel. 15, Created)  
DT 01-AUG-1990 (Rel. 15, Last sequence update)  
DT 01-NOV-1995 (Rel. 32, Last annotation update)  
DE PROTEIN KINASE DC1A (EC 2.7.1.37).  
DC1 OR PRA-C2.  
GN Drosophila melanogaster (Fruit fly).  
OS Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CANTON-S;  
RX MEDLINE; 89107990.  
RA KALDERON D., RUBIN G.M.;  
RT "Isolation and characterization of Drosophila cAMP-dependent protein  
RT kinase genes".  
RL Genes Dev. 2:1539-1556(1988).  
CC -1- CATALYTIC ACTIVITY: ATP + A PROTEIN = ADP + A PHOSPHOPROTEIN.  
CC -1- ALTERNATIVE PRODUCTS: DC1A AND DC1B ARE PRODUCED BY THE  
CC ALTERNATIVE SPLICING OF THE SAME GENE.  
CC -1- TISSUE SPECIFICITY: MORE ABUNDANT IN ADULT BODY THAN ADULT HEAD.  
CC -1- DEVELOPMENTAL STAGE: IN EMBRYONS, PUPAE AND ADULTS.  
CC -1- SIMILARITY: STRONG TO CAMP-DEPENDENT PROTEIN KINASE CATALYTIC  
CC SUBUNIT (DC0).  
CC  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; X16950; CAA34833.1; -  
CC PIR; D31751; D31751.  
CC HSSP; P05132; LCTP.  
CC FLYBASE; Fgn0000274; Pka-C2.  
CC PROSITE; P500107; PROTEIN\_KINASE\_ATP; 1.  
CC PROSITE; P500108; PROTEIN\_KINASE\_ST; 1.  
CC PROSITE; P500111; PROTEIN\_KINASE\_DOM; 1.  
CC PFAM; PF00069; pkinase; 1.  
CC PFAM; PF00433; pkinase\_C; 1.  
CC  
CC Transferase; Serine/threonine-protein kinase; ATP-binding;  
CC Alternative splicing.  
CC FT DOMAIN 45 301 PROTEIN KINASE.  
CC FT NP\_BIND 51 59 ATP (BY SIMILARITY).  
CC FT BINDING 74 74 ATP (BY SIMILARITY).  
CC FT ACT\_SITE 168 168 BY SIMILARITY.  
CC SQ SEQUENCE 354 AA; 41468 MW; 490E4156 CRC32;  
  
Query Match 91.1%; Score 41; DB 1; Length 354;  
Best Local Similarity 41.7%; Pred. No. 1.08e+01;  
Matches 5; Conservative 0; Mismatches 7; Indels 0; Gaps 0;  
  
Db 239 PFAIHDRVDVILM 250  
QY 2 PXXXXXXAVILM 13  
  
RESULT 15  
ID AROC\_ECOLI STANDARD; PRT; 360 AA.  
AC P12008; P78193;  
DT 01-OCT-1989 (Rel. 12, Created)  
DT 01-MAY-1991 (Rel. 18, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE CHORISMATE SYNTHASE (EC 4.6.1.4) (5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE  
DE DE PHOSPHOLYASE).  
GN AROC.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Escherichia.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 90218018.  
RA CHARLES I.G., LAMB H.K., PICKARD D., DOUGAN G., HAWKINS A.R.;  
RT "Isolation, characterization and nucleotide sequences of the aroc  
RT genes encoding chorismate synthase from Salmonella typhi and  
RT Escherichia coli".  
RL J. Gen. Microbiol. 136:353-358(1990).  
RN [2]  
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
RC STRAIN=K12;  
RX MEDLINE; 88293429.  
RA WHITE P.J., MILLAR G., COGGINS J.R.;  
RT "The overexpression, purification and complete amino acid sequence of  
RT chorismate synthase from Escherichia coli K12 and its comparison with  
RT the enzyme from Neurospora crassa".  
RL Biochem. J. 251:313-322(1988).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12 / MG1655;  
RX MEDLINE; 97426617.  
RA BLATTNER F.R., PLUNKETT G. III, BLOCH C.A., PERNA N.T., BURLAND V.,  
RA RILEY M., COLLADO-VIDES J., GLASNER F.D., RODE C.K., MAYHEW G.F.,  
RA GREGOR J., DAVIS N.W., KIRKPATRICK H.A., GORDEN M.A., ROSE D.J.,  
RA MAU B., SHAO Y.;  
RT "The complete genome sequence of Escherichia coli K-12".  
RL Science 277:1453-1474(1997).  
RN [4]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12;  
RA AIBA H., BABA T., FUJITA K., HAYASHI K., HONJO A., HORIUCHI T.,  
RA IKEMOTO K., INADA T., ISONO K., ISONO S., ITOH T., KANAI K.,



DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE TRANSFORMING PROTEIN RHOC.  
GN ARHC OR ARH9 OR RHOC.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;  
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 95173444.  
RA SEGAGE F., CLAUDIO E., WROBEL K., RAMOS S., LAZO P.S.;  
RT "Isolation of nine gene sequences induced by silica in murine  
RT macrophages".  
RT J. Immunol. 154:2384-2392(1995).  
CC -!- FUNCTION: REGULATES A SIGNAL TRANSDUCTION PATHWAY LINKING PLASMA  
CC MEMBRANE RECEPTORS TO THE ASSEMBLY OF FOCAL ADHESIONS AND ACTIN  
CC STRESS FIBERS.  
CC -!- SIMILARITY: BELONGS TO THE RHO FAMILY IN THE RAS SUPERFAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; X80638; CAA35682.1; -.  
DR HSP; P06749; 1A2B.  
DR MGD; MG1:106028; ARHC.  
DR PFAM; PF00071; ras; 1.  
KW Proto-oncogene; GTP-binding; Prenylation; Lipoprotein.  
FT NP\_BIND 12 19 GTP (BY SIMILARITY).  
FT NP\_BIND 59 63 GTP (BY SIMILARITY).  
FT NP\_BIND 117 120 GTP (BY SIMILARITY).  
FT DOMAIN 34 42 EFFECTOR REGION (POTENTIAL).  
FT LIPID 190 190 GERANYL-GERANYL (BY SIMILARITY).  
FT SEQUENCE 193 AA; 22005 MW; 62B8C8AE CRC32;  
SQ  
Query Match 91.1%; Score 41; DB 1; Length 193;  
Best Local Similarity 41.7%; Pred. No. 1.08e+01;  
Matches 5; Conservative 0; Mismatches 7; Indels 0; Gaps 0;  
DB 71 PLSYPDTDVILM 82  
QY 2 PXXXXXXAVILM 13  
RESULT 12  
ID RHOA\_CANFA STANDARD; PRT; 193 AA.  
AC F24406;  
DT 01-MAR-1992 (Rel. 21, Created)  
DT 01-MAR-1992 (Rel. 21, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE TRANSFORMING PROTEIN RHOA (RH01).  
GN ARHA OR ARH12 OR RHOA OR RH01.  
OS Canis familiaris (Dog).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;  
OC Eutheria; Carnivora; Fissipedia; Canidae; Canis.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN-COCKER SPANIEL;  
RX MEDLINE; 91061765.  
RA CHAVIER P., VINGRON M., SANDER C., SIMONS K., ZERIAL M.;  
RT "Molecular cloning of YP71/SEC4-related cDNAs from an epithelial cell  
RT line".  
RT Mol. Cell. Biol. 10:6578-6585(1990).  
RL Mol. Cell. Biol. 10:6578-6585(1990).  
CC -!- FUNCTION: REGULATES A SIGNAL TRANSDUCTION PATHWAY LINKING PLASMA  
CC MEMBRANE RECEPTORS TO THE ASSEMBLY OF FOCAL ADHESIONS AND ACTIN  
CC STRESS FIBERS.  
CC -!- PTM: SUBSTRATE FOR BOTULINUM ADP-RIBOSYLTRANSFERASE.  
CC -!- SIMILARITY: BELONGS TO THE RHO FAMILY IN THE RAS SUPERFAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; X56391; CAA39802.1; -.  
DR PIR; H36364; H36364.  
DR HSP; P06749; 1A2B.  
DR PFAM; PF00071; ras; 1.  
KW Proto-oncogene; GTP-binding; Prenylation; Lipoprotein.  
FT PROPEP 191 193 REMOVED IN MATURE FORM (BY SIMILARITY).  
FT NP\_BIND 12 19 GTP (BY SIMILARITY).  
FT NP\_BIND 59 63 GTP (BY SIMILARITY).  
FT NP\_BIND 117 120 GTP (BY SIMILARITY).  
FT DOMAIN 34 42 EFFECTOR REGION (POTENTIAL).  
FT LIPID 190 190 GERANYL-GERANYL (BY SIMILARITY).  
FT SEQUENCE 193 AA; 21740 MW; 80F8D701 CRC32;  
SQ  
Query Match 91.1%; Score 41; DB 1; Length 193;  
Best Local Similarity 41.7%; Pred. No. 1.08e+01;  
Matches 5; Conservative 0; Mismatches 7; Indels 0; Gaps 0;  
DB 71 PLSYPDTDVILM 82  
QY 2 PXXXXXXAVILM 13  
RESULT 13  
ID RHOB\_HUMAN STANDARD; PRT; 196 AA.  
AC P01121;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 01-AUG-1988 (Rel. 08, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE TRANSFORMING PROTEIN RHOB (H6).  
GN ARHB OR ARH6 OR RHOB.  
OS Homo sapiens (Human), Mus musculus (Mouse), and  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Primates; Catarrhini; Homnidae; Homo.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX SPECIES-HUMAN;  
RX MEDLINE; 88203210.  
RA CHARDIN P., MADAULE P., TAVITIAN A.;  
RT "Coding sequence of human rho cDNAs clone 6 and clone 9".  
RL Nucleic Acids Res. 16:2717-2717(1988).  
RN [2]  
RP SEQUENCE OF 29-196 FROM N.A.  
RC SPECIES-HUMAN;  
RX MEDLINE; 85201682.  
RA MADAULE P., AXEL R.;  
RT "A novel ras-related gene family".  
RL Cell 41:31-40(1985).  
RN [3]  
RP SEQUENCE FROM N.A.  
RX SPECIES-RAT;  
RX MEDLINE; 91260717.  
RA JAENER D., HUNTER T.;  
RT "The ras-related gene rhoB is an immediate-early gene inducible by v-  
RT fps, epidermal growth factor, and platelet-derived growth factor in  
RT rat fibroblasts".  
RL Mol. Cell. Biol. 11:3682-3690(1991).  
RN [4]  
RP SEQUENCE FROM N.A.  
RX SPECIES-MOUSE;  
RX MEDLINE; 96428574.  
RA NAKAMURA T., ASANO M., SHINDO-OKADA N., NISHIMURA S., MONDEN Y.;  
RT "Cloning of the rhoB gene from the mouse genome and characterization  
RT of its promoter region".  
RL Biochem. Biophys. Res. Commun. 226:688-694(1996).

```

Query Match          91.18; Score 41; DB 1; Length 193;
Best Local Similarity 41.7%; Pred. No. 1.08e+01;
Matches          5; Conservative          0; Mismatches 7; Indels          0; Gaps          0;

Db      71  PLSYPDTDVILM 82
      |
      2  PXXXXXXAVILM 13
      |
      |
      |

RESULT 10
ID RHOC_HUMAN STANDARD; PRT; 193 AA.
AC P08134;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE TRANSFORMING PROTEIN RHOC (H9).
GN ARHC OR ARH9 OR RHOC.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Homnidae; Homo.
RN [1]
RN SEQUENCE FROM N.A.
RP RP
RX MEDLINE; 88203210.
RA CHARDIN P., MADAULE P., TAVITIAN A.;
RT "Coding sequence of human rho cDNAs clone 6 and clone 9.";
RL Nucleic Acids Res. 16:2717-2717(1988).
RN [2]
RN SEQUENCE FROM N.A.
RP RP
RC TISSUE=RETINA;
RA FAGAN K.P., OLIVEIRA L., PITTLE S.J.;
RL Submitted (OCT-1993) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: REGULATES A SIGNAL TRANSDUCTION PATHWAY LINKING PLASMA
CC MEMBRANE RECEPTORS TO THE ASSEMBLY OF FOCAL ADHESIONS AND ACTIN
CC STRESS FIBERS.
CC -1- SIMILARITY: BELONGS TO THE RHO FAMILY IN THE RAS SUPERFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X06821; CAA29969.1; -
CC EMBL; L25081; AAC33179.1; -
CC PIR; S01029; TVHURC.
CC DR HSSP; P06749; 1A2B.
CC DR MIM; 165380; -
CC DR PFAM; PF00071; ras; 1.
KW Proto-oncogene, GTP-binding; Prenylation; Lipoprotein.
FT NP_BIND 12 19 GTP (BY SIMILARITY).
FT FT NP_BIND 59 63 GTP (BY SIMILARITY).
FT FT NP_BIND 117 120 GTP (BY SIMILARITY).
FT FT DOMAIN 34 42 EFFECTOR REGION (POTENTIAL).
FT FT LIPID 190 190 GERANYL-GERANYL (BY SIMILARITY).
SQ SEQUENCE 193 AA; 22006 MW; F803713B CRC32;

Query Match          91.18; Score 41; DB 1; Length 193;
Best Local Similarity 41.7%; Pred. No. 1.08e+01;
Matches          5; Conservative          0; Mismatches 7; Indels          0; Gaps          0;

Db      71  PLSYPDTDVILM 82
      |
      2  PXXXXXXAVILM 13
      |
      |
      |

RESULT 11
ID RHOC_MOUSE STANDARD; PRT; 193 AA.
AC Q62159;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)

```

```
FT LIPID 189 189 GERANYL-GERANYL (BY SIMILARITY).
SQ SEQUENCE 192 AA; 21635 MW; F52C3088 CRC32;

Query Match 91.1%; Score 41; DB 1; Length 192;
Best Local Similarity 41.7%; Pred. No. 1.08e+01;
Matches 5; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Db 71 PLSYPDTVDILM 82
QY 1
2 PXXXXXXAVILM 13

RESULT 7
ID RHO1_DROME STANDARD; PRT; 192 AA.
AC P48148;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE RAS-LIKE GTP-BINDING PROTEIN RHO1.
GN RHO1.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 95137009.
RA HARTIHARAN I.K., HU K.-Q., ASHA H., QUINTANILLA A., EZZELL R.M.,
RA SETTLEMAN J.;
RT melanogaster: overexpressing Rho1 in retinal cells causes a late
RT developmental defect.;
RL EMBO J. 14:292-302(1995).
CC -1- SIMILARITY: BELONGS TO THE RHO FAMILY IN THE RAS SUPERFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: M10078; AAA27776.1;
CC PIR: A01373; TVGNAC.
CC HSSP; P06749; IAZB.
CC PFAM; PF00071; ras; 1.
CC GTP-binding; Prenylation; Lipoprotein.
FT NP_BIND 12 19 GTP (BY SIMILARITY).
FT NP_BIND 59 63 GTP (BY SIMILARITY).
FT NP_BIND 117 120 GTP (BY SIMILARITY).
FT DOMAIN 34 42 EFFECTOR REGION (POTENTIAL).
FT LIPID 189 189 GERANYL-GERANYL (BY SIMILARITY).
SQ SEQUENCE 192 AA; 21723 MW; 9355C94D CRC32;

Query Match 91.1%; Score 41; DB 1; Length 192;
Best Local Similarity 41.7%; Pred. No. 1.08e+01;
Matches 5; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Db 71 PLSYPDTVDILM 82
QY 1
2 PXXXXXXAVILM 13

RESULT 8
ID RHO.APLCA STANDARD; PRT; 192 AA.
AC P01122;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 01-OCT-1994 (Rel. 30, Last annotation update)
DE RAS-LIKE GTP-BINDING PROTEIN RHO.
GN RHO.
OS Aplysia californica (California sea hare).
OC Eukaryota; Metazoa; Mollusca; Gastropoda; Opisthobranchia; Anaspidia;
OC Aplysiidae; Aplysia.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 85201682.
RA MADAULE P., AXEL R.;
RT "A novel ras-related gene family.";
RL Cell 41:31-40(1985).
CC -1- SIMILARITY: BELONGS TO THE RHO FAMILY IN THE RAS SUPERFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: M10078; AAA27776.1;
CC PIR: A01373; TVGNAC.
CC HSSP; P06749; IAZB.
CC PFAM; PF00071; ras; 1.
CC GTP-binding; Prenylation; Lipoprotein.
FT NP_BIND 12 19 GTP (BY SIMILARITY).
FT NP_BIND 59 63 GTP (BY SIMILARITY).
FT NP_BIND 117 120 GTP (BY SIMILARITY).
FT DOMAIN 34 42 EFFECTOR REGION (POTENTIAL).
FT LIPID 189 189 GERANYL-GERANYL (BY SIMILARITY).
SQ SEQUENCE 192 AA; 21661 MW; BEE6C230 CRC32;

Query Match 91.1%; Score 41; DB 1; Length 192;
Best Local Similarity 41.7%; Pred. No. 1.08e+01;
Matches 5; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Db 71 PLSYPDTVDILM 82
QY 1
2 PXXXXXXAVILM 13

RESULT 9
ID RHOA_HUMAN STANDARD; PRT; 193 AA.
AC P06749;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE TRANSFORMING PROTEIN RHOA (H12).
GN ARHA OR ARH12 OR RHOA OR RHO12.
OS Homo sapiens (Human), and Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 87146500.
RA YERAMIAN P., CHARDIN P., MADAULE P., TAVITTIAN A.;
RT "Nucleotide sequence of human rho cDNA clone 12.";
RL Nucleic Acids Res. 15:1869-1869(1987).
CC [2]
CC SEQUENCE FROM N.A.
CC SPECIES-HUMAN;
CC SPECIES-HUMAN;
CC FAGAN K.P., OLIVEIRA L., PITTLER S.J.;
CC Submitted (OCT-1993) to the EMBL/GenBank/DBJ databases.
CC [3]
CC SEQUENCE OF 137-193 FROM N.A.
CC SPECIES-HUMAN;
CC MEDLINE; 92210561.
CC MOSCOW J.A., MORROW C.S., HE R., MULLENBACH G.T., CORAN K.H.;
CC "Structure and function of the 5'-flanking sequence of the human
CC cytosolic selenium-dependent glutathione peroxidase gene (hgp1).";
CC J. Biol. Chem. 267:5949-5958(1992).
CC [4]
CC SEQUENCE FROM N.A.
CC SPECIES-BOVINE;
```

```

-----
RESULT 4
ID EF1B_ARCFU STANDARD; PRT; 88 AA.
AC Q29681;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE ELONGATION FACTOR 1-BETA (EF-1-BETA).
GN AF0574.
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobales; Archaeoglobaceae;
OC Archaeoglobus.
CC [1]
CC SEQUENCE FROM N.A.
CC STRAIN-VC-16 / DSM 4304 / ATCC 49558;
CC MEDLINE; 98049343.
CC KLENK H.-P., CLAYTON R.A., TOMB J.-F., WHITE O., NELSON K.F.,
CC KETCHUM K.A., DODSON R.J., GWINN M., HICKEY E.K., PETERSON J.D.,
CC RICHARDSON D.L., KERLAJAVE A.R., GRAHAM D.E., KYRPIDES N.C.,
CC FLEISCHMANN R.D., QUACKENBUSH J., LEE N.H., SUTTON G.G., GILL S.,
CC KIRKNESS E.F., DOUGHERTY B.A., MCKENNEY K., ADAMS M.D., LOFTUS B.,
CC PETERSON S., REICH C.I., MCNEIL L.K., BADGER J.H., GLODEK A., ZHOU L.,
CC COTTON M.D., SPRIGGS T., WEIDMAN J.F., McDONALD L., UTTERBACK T.,
CC SADOW P.W., D'ANDREA K.P., BOWMAN C., FUJII C., GARLAND S.A.,
CC MASON T.M., OLSEN G.J., FRASER C.M., SMITH H.O., WOESSE C.R.,
CC VENTER J.C.;
CC "The complete genome sequence of the hyperthermophilic, sulphate-
CC reducing archaeon Archaeoglobus fulgidus."
CC Nature 390:364-370(1997).
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; M38396; AAA49225.1;
CC PIR; G38625; G38625.
CC HSSP; P06749; 1A2B.
CC PFAM; PF00071; ras; 1.
CC KW GTP-Binding; Prenylation; Lipoprotein.
CC FT NP_BIND 12 19 GTP (BY SIMILARITY).
CC FT NP_BIND 59 63 GTP (BY SIMILARITY).
CC FT NP_BIND 117 120 GTP (BY SIMILARITY).
CC FT DOMAIN 34 42 EFFECTOR REGION (POTENTIAL).
CC FT LIPID 189 189 GERANYL-GERANYL (BY SIMILARITY).
CC SQ SEQUENCE 192 AA; 21480 MW; 5B24A92B CRC32;
CC
CC Query Match 91.1%; Score 41; DB 1; Length 192;
CC Best Local Similarity 41.7%; Pred. No. 1.08e+01;
CC Matches 5; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
CC
CC Db 71 PLSYPDTDVILM 82
CC | | | | |
CC QY 2 PXXXXXXAVILM 13
-----
RESULT 6
ID RHOA_CAEEL STANDARD; PRT; 192 AA.
AC Q22038;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE RAS-LIKE GTP-BINDING PROTEIN RHOA.
GN RHOA.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
OC Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
CC [1]
CC SEQUENCE FROM N.A.
CC STRAIN-BRISTOL N2;
CC RX MEDLINE; 95096090.
CC RA CHEN W., LIM L.;
CC RT "The Caenorhabditis elegans small GTP-binding protein RhoA is
CC enriched in the nerve ring and sensory neurons during larval
CC development."
CC J. Biol. Chem. 269:32394-32404(1994).
CC CC -|- SUBCELLULAR LOCATION: ASSOCIATED WITH THE MEMBRANE AND THE
CC CYTOSKELETON THROUGHOUT DEVELOPMENT.
CC -|- DEVELOPMENTAL STAGE: UBIQUITOUS EXPRESSION THROUGHOUT DEVELOPMENT
CC WITH A PARTICULAR ENRICHMENT AT LARVAL STAGES IN THE PHARYNGEAL
CC NERVE RING AND AT THE TIP OF THE HEAD CONTAINING CHEMOSENSORY AND
CC MECHANOSENSORY NEURONS.
CC -|- SIMILARITY: BELONGS TO THE RHO FAMILY IN THE RAS SUPERFAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; L36965; AAC37216.1;
CC HSSP; P06749; 1A2B.
CC PFAM; PF00071; ras; 1.
CC KW GTP-Binding; Prenylation; Lipoprotein.
CC FT NP_BIND 12 19 GTP (BY SIMILARITY).
CC FT NP_BIND 59 63 GTP (BY SIMILARITY).
CC FT NP_BIND 117 120 GTP (BY SIMILARITY).
CC FT DOMAIN 34 42 EFFECTOR REGION (POTENTIAL).
CC
CC Query Match 91.1%; Score 41; DB 1; Length 88;
CC Best Local Similarity 33.3%; Pred. No. 1.08e+01;
CC Matches 4; Conservative 2; Mismatches 6; Indels 0; Gaps 0;
CC
CC Db 43 PIAFGLKAVILM 54
CC | | | | |
CC QY 2 PXXXXXXAVILM 13
-----
RESULT 5
ID RHO_DISOM STANDARD; PRT; 192 AA.
AC P22122;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 01-OCT-1993 (Rel. 27, Last annotation update)
DE RAS-LIKE GTP-BINDING PROTEIN O-RHO.
OS Discopige omata (Electric ray).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
OC Elasmobranchii; Rajiformes; Torpedinoidei; Narciniidae; Discopige.
CC [1]
CC SEQUENCE FROM N.A.
CC TISSUE-ELECTRIC LOBE;
CC RX MEDLINE; 91115900.
CC RA NGSEE J.K., ELFERINK L.A., SCHELLER R.H.;
CC RT "A family of ras-like GTP-binding proteins expressed in electromotor
CC neurons."
CC J. Biol. Chem. 266:2675-2680(1991).
CC -|- SIMILARITY: BELONGS TO THE RHO FAMILY IN THE RAS SUPERFAMILY.
-----

```



\*\*\*\*\*  
W O R L D  
(TM)  
\*\*\*\*\*  
Release 3.1A John F. Collins, BioComputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd  
MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Sat Apr 15 01:20:26 2000; Maspar time 3.20 Seconds  
Tabular output not generated. 121.400 Million cell updates/sec

Title: >US-08-452-843-22  
Description: (1-13) from US08452843.pap  
Perfect Score: 45  
Sequence: 1 XPXXXXXXAVILM 13  
Scoring table: PAM 150  
Gap 15  
Searched: 82229 seqs, 29864866 residues  
Post-processing: Minimum Match 0%  
Listing first 45 summaries  
Database: swiss-prot38  
1:swissprot  
Statistics: Mean 20.915; Variance 19.563; scale 1.069  
Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	45	100.0	436	1 YFCY_ECOLI	PROBABLE 3-KETOACYL-CO	1.02e+00
2	43	95.6	681	1 SNAV_SALTY	SECRETION SYSTEM APPAR	3.38e+00
3	43	95.6	1034	1 ACRF_ECOLI	ACRYLAMINE RESISTANCE	3.38e+00
4	41	91.1	88	1 EFIB_ARCFU	ELONGATION FACTOR 1-BE	1.08e+01
5	41	91.1	88	1 RHO_DISOM	RAS-LIKE GTP-BINDING P	1.08e+01
6	41	91.1	192	1 RHO1_DISEL	RAS-LIKE GTP-BINDING P	1.08e+01
7	41	91.1	192	1 RHO1_DROME	RAS-LIKE GTP-BINDING P	1.08e+01
8	41	91.1	192	1 RHO1_APLCA	RAS-LIKE GTP-BINDING P	1.08e+01
9	41	91.1	193	1 RHOA_HUMAN	TRANSFORMING PROTEIN R	1.08e+01
10	41	91.1	193	1 RHOA_HUMAN	TRANSFORMING PROTEIN R	1.08e+01
11	41	91.1	193	1 RHOC_MOUSE	TRANSFORMING PROTEIN R	1.08e+01
12	41	91.1	193	1 RHOA_CANFA	TRANSFORMING PROTEIN R	1.08e+01
13	41	91.1	196	1 RHOB_HUMAN	TRANSFORMING PROTEIN R	1.08e+01
14	41	91.1	354	1 KDCA_DROME	PROTEIN KINASE DC1A (E	1.08e+01
15	41	91.1	360	1 AROC_ECOLI	CHORISMATE SYNTHASE (E	1.08e+01
16	41	91.1	360	1 AROC_VIBAN	CHORISMATE SYNTHASE (E	1.08e+01
17	41	91.1	372	1 AROC_VIBAN	CHORISMATE SYNTHASE (E	1.08e+01
18	41	91.1	376	1 KDCE_DROME	PROTEIN KINASE DC1B (E	1.08e+01
19	41	91.1	1090	1 NIT4_NEUCR	NITROGEN ASSIMILATION	1.08e+01
20	40	88.9	103	1 VAS3_VACCV	PROTEIN A53.	1.89e+01
21	40	88.9	103	1 VAS3_VACCV	PROTEIN A53.	1.89e+01
22	40	88.9	375	1 GM12_SCHPO	ALPHA-1,2-GALACTOSYLTR	1.89e+01
23	40	88.9	493	1 YTH2_RHOSN	HYPOTHETICAL 53.0 KD G	1.89e+01

24	40	88.9	782	1 YAKB_SCHPO	PUTATIVE 89.3 KD TRANS	1.89e+01
25	40	88.9	971	1 VP2_EHDV1	OUTER CAPSID PROTEIN V	1.89e+01
26	40	88.9	1365	1 KRE5_YEAST	KILLER TOXIN-RESISTANC	1.89e+01
27	39	86.7	102	1 CH10_CHLPN	10 KD CHAPERONIN (PROT	3.28e+01
28	39	86.7	191	1 RHOB_HUMAN	RHO-RELATED GTP-BINDIN	3.28e+01
29	39	86.7	326	1 VS09_ROTSH	GLYCOPROTEIN VP7 (SERO	3.28e+01
30	39	86.7	326	1 VS09_ROTSH	GLYCOPROTEIN VP7 (SERO	3.28e+01
31	39	86.7	326	1 VS09_ROTSH	GLYCOPROTEIN VP7 (SERO	3.28e+01
32	39	86.7	326	1 VS09_ROTSH	GLYCOPROTEIN VP7 (SERO	3.28e+01
33	39	86.7	326	1 VS09_ROTSH	GLYCOPROTEIN VP7 (SERO	3.28e+01
34	39	86.7	326	1 VS09_ROTSH	GLYCOPROTEIN VP7 (SERO	3.28e+01
35	39	86.7	326	1 VS09_ROTSH	GLYCOPROTEIN VP7 (SERO	3.28e+01
36	39	86.7	326	1 VS09_ROTSH	GLYCOPROTEIN VP7 (SERO	3.28e+01
37	39	86.7	326	1 VS09_ROTSH	GLYCOPROTEIN VP7 (SERO	3.28e+01
38	39	86.7	630	1 YH13_SCHPO	HYPOTHETICAL 70.6 KD P	3.28e+01
39	39	86.7	1037	1 YH13_SCHPO	HYPOTHETICAL 111.5 KD	3.28e+01
40	39	86.7	1040	1 AXOL_ECOLI	AXONIN-1 PRECURSOR (AX	3.28e+01
41	39	86.7	1836	1 CIN4_HUMAN	SODIUM CHANNEL PROTEIN	3.28e+01
42	38	84.4	349	1 ADH1_ASPEL	ALCOHOL DEHYDROGENASE	5.62e+01
43	38	84.4	417	1 PGK1_RHINI	PHOSPHOGLYCERATE KINAS	5.62e+01
44	38	84.4	751	1 ABP_HUMAN	AMILORIDE-SENSITIVE AM	5.62e+01
45	38	84.4	3176	1 CA35_HUMAN	COLLAGEN ALPHA 3(VI) C	5.62e+01

ALIGNMENTS

RESULT 1  
ID YFCY\_ECOLI STANDARD; PRT; 436 AA.  
AC P76503;  
DT 15-JUL-1999 (Rel. 38, Created)  
DT 15-JUL-1999 (Rel. 38, Last sequence update)  
DE PROBABLE 3-KETOACYL-COA THIOLASE (EC 2.3.1.16) (ACETYL-COA  
DE ACYLTRANSFERASE) (BETA-KETOTHIOLASE).  
GN YFCI.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Escherichia.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-K12 / MG1655;  
RX MEDLINE; 97426617.  
RA BLATTNER F.R., PLUNKETT G. III, BLOCH C.A., PERNA N.T., BURLAND V.,  
RA RILEY M., COLLADO-VIDES J., GLASNER F.D., RODE C.K., MAYHEW G.F.,  
RA GREGOR J., DAVIS N.W., KIRKPATRICK H.A., GOEDEN M.A., ROSE D.J.,  
RA MAU B., SHAO Y.;  
RT "The complete genome sequence of Escherichia coli K-12.";  
RL Science 277:1453-1474(1997).  
CC -|- CATALYTIC ACTIVITY: ACYL-COA + ACETYL-COA -> COA + 3-OXOACYL-COA.  
CC -|- SUBCELLULAR LOCATION: CYTOPLASMIC (POTENTIAL).  
CC -|- SIMILARITY: BELONGS TO THE THIOLASE FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC -----  
CC EMBL; AE000322; AAC75402.1;  
CC ECOGENE; EG14128; YFCY.  
DR PROSITE; PS00098; THIOLASE\_1; 1.  
DR PROSITE; PS00737; THIOLASE\_2; 1.  
DR PROSITE; PS00099; THIOLASE\_3; 1.  
DR PFAM; PF0108; thiolase; 1.  
KW Hypothetical protein; Transferase; Acyltransferase.  
FT ACT SITE 99 99 SUBSTRATE BINDING (BY SIMILARITY).  
FT ACT SITE 422 422 BASE (BY SIMILARITY).  
SQ SEQUENCE 436 AA; 46530 MW; 8BA3CBD4 CRC32;  
Query Match 100.0%; Score 45; DB 1; Length 436;  
Best Local Similarity 50.0%; Pred. NO. 1.02e+00;

Matches 4; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

Db 193 PWNWLDVSVIMM 204  
QY 2 PXXXXXXAVILM 13

RESULT 15

ID R82534 standard; Protein; 37 AA.  
AC R82534;  
DT 16-APR-1996 (first entry)  
DE IE beta chain fragment.  
KW Major histocompatibility complex; MHC; T-cell receptor; TCR;  
KW autoimmune disease; immunodeficiency disease; immune response;  
KW immunoproliferation disease; graft-host rejection; therapy.  
OS Synthetic.  
FH Key Location/Qualifiers  
FT misc\_feature 1..27 "IE beta leader sequence"  
FT misc\_feature 28..37  
FT /note= "region of IE beta chain betal domain"  
PN WO9523814-A1.  
PD 08-SEP-1995.  
PE 03-MAR-1995; U02689  
PR 04-MAR-1994; US-207481.  
PA (NAJE-) NAT JEWISH CENT IMMUNOLOGY & RESPIRATORY.  
PI Kappler JW, Marrack P;  
DR WPI: 95-320543/41.  
DR N-PSDB: T04265.  
PT Peptide-MHC complex comprising antigenic peptide, linker and MHC  
PT segment - useful as reagents for the treatment of diseases including  
PT auto-immune diseases, immuno-stimulatory diseases or graft-host  
PT rejection  
PS Example 1; Page 56; 94pp; English.  
CC This sequence represents the start of the IE beta chain. The DNA  
CC encoding this sequence was used in the construction of a hybrid beta  
CC chain containing the moth cyto c peptide (MCC) (see R04266). The hybrid  
CC sequence was then used in the construction of a hybrid IE alpha beta  
CC dimer. The encoded protein (IE K/d-MCC) was found to be more stable than  
CC the IE alpha beta dimer. The stability was increased even further by the  
CC addition of a MHC groove specific binding peptide (e.g. see R82527,  
CC R82528 and R82531). These complexes may be used to regulate an immune  
CC response. The complexes are capable of being recognised by a TCR alone  
CC or in combination with additional MHC proteins. These complexes are  
CC useful for therapeutic purposes and experimental purposes. They can also  
CC be used as reagents for the treatment of diseases including autoimmune  
CC diseases, immunodeficiency diseases, immunoproliferation diseases, and  
CC graft-host rejection.  
SQ Sequence 37 AA;

Query Match 84.4%; Score 38; DB 1; Length 37;  
Best Local Similarity 41.7%; Pred. No. 5.07e+02;  
Matches 5; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

Db 5 PRVPCVAVILL 16  
QY 2 PXXXXXXAVILM 13

Search completed: Sat Apr 15 01:19:35 2000  
Job time : 38 secs.

```
PN W08505122-A.
PD 21-NOV-1985.
PF 29-APR-1985; AU00096.
PR 27-APR-1984; AU-004733.
PR 01-JAN-1985; AU-042970.
PR 29-APR-1985; WO-AU00096.
PA (UYME-) UNIV OF MELBOURNE.
PI Holmes I, Dyal-Smith ML;
DR WPI: 86-028178/04.
DR N-PSDB: N60458.
PT RNA gene segment coding for outer capsid glycoprotein of
PT rotavirus - useful in expression of antigenic viral proteins by
PT bacteria for use in vaccines and diagnostic prods.
PS Claim 6; Fig 4; 24pp; English.
CC The rotavirus segment 8 encoding sequence was isolated from the Hu/
CC Australia/5/11. The product and fragments comprise at least part of
CC the major outer capsid glycoprotein, they may be expressed from a
CC transformed host and are useful as antigens for vaccination and
CC diagnosis of the rotavirus.
SQ Sequence 326 AA;

Query Match 86.7%; Score 39; DB 1; Length 326;
Best Local Similarity 33.3%; Pred. No. 3.92e+02;
Matches 4; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

Db 131 POLYCDYNVVLM 142
QY 2 PXXXXXXAVILM 13

RESULT 12
ID W44119 standard; Protein; 1479 AA.
AC W44119;
DT 12-MAY-1998 (first entry)
DE Human type C lectin.
KW Murine; human; type C lectin; E-selectin; competitive inhibitor;
KW mouse; molecular marker.
OS Homo sapiens.
PN W09740154-A1.
PD 30-OCT-1997.
PF 17-APR-1997; U06347.
PR 24-APR-1996; U0637021.
PA (GETH) GENENTECH INC.
PI Lasky LA, Wu K;
DR WPI: 97-535838/49.
DR N-PSDB: V02186.
PT Human and mouse type C lectin(s) - useful as competitive inhibitor
PT of lectin activity and as molecular markers for tissues that express
PT them
PS Claim 1; Page 51-56; 97pp; English.
CC The present sequence represents a novel type C lectin. Type C lectins
CC can be used to identify and purify their native ligands and compete
CC with them for their binding, especially useful as competitive
CC inhibitors of the biological activity of native type C lectins.
CC This makes type C lectins useful as molecular markers for tissues
CC in which they are expressed. The nucleic acid, encoding type C lectin,
CC can be used to identify other type C lectin coding sequences.
SQ Sequence 1479 AA;

Query Match 86.7%; Score 39; DB 1; Length 1479;
Best Local Similarity 33.3%; Pred. No. 3.92e+02;
Matches 4; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

Db 1409 PENPAALVVVLM 1420
QY 2 PXXXXXXAVILM 13

RESULT 13
ID W69361 standard; Protein; 1978 AA.
AC W69361;
DT 01-DEC-1998 (first entry)
DE Tetrodotoxin-sensitive sodium channel PN4.

KW Tetrodotoxin-sensitive sodium channel; rat; PN4 sodium channel; stroke;
KW nervous system disorder; epilepsy; brain injury; diabetic neuropathy;
KW AIDS-associated neuropathy; therapy.
OS Rattus sp.
PN W09838302-A2.
PD 03-SEP-1998.
PF 20-FEB-1998; E00997.
PR 26-FEB-1997; US-039447.
PA (HOFF) HOFFMANN LA ROCHE & CO AG F.
PI Delgado SG, Dietrich PS, Fish LM, Herman RC, Sangameswaran L;
DR WPI: 98-481204/41.
DR N-PSDB: V58420.
PT New rat tetrodotoxin-sensitive sodium channel alpha subunit and DNA
PT - for detecting inhibitors which alleviate pain, and treating
PT nervous system disorders, e.g. epilepsy, stroke, diabetic and AIDS
PT neuropathy
PS Claim 16; Page 36-43; 87pp; English.
CC This sequence is encoded by the rat PN4 sodium channel cDNA clone of the
CC invention. The DNA sequence was isolated from a peripheral nerve from
CC a rat dorsal ganglia. The PN4 sodium channel sequences are
CC tetrodotoxin-sensitive sodium channels. The protein is used in assays for
CC detecting inhibitors of tetrodotoxin-sensitive sodium channels, which
CC alleviate pain. The probes can be used to detect and isolate the DNA or
CC protein in tissues. The antibodies can also be used to isolate the
CC protein. The protein is used as a therapeutic target for compounds to
CC treat disorders of the nervous system, such as epilepsy, stroke and brain
CC injury, diabetic neuropathy, and AIDS-associated neuropathy, etc.
SQ Sequence 1978 AA;

Query Match 86.7%; Score 39; DB 1; Length 1978;
Best Local Similarity 33.3%; Pred. No. 3.92e+02;
Matches 4; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

Db 193 PNWMLDFSVIMM 204
QY 2 PXXXXXXAVILM 13

RESULT 14
ID W69362 standard; Protein; 1988 AA.
AC W69362;
DT 01-DEC-1998 (first entry)
DE Tetrodotoxin-sensitive sodium channel PN4a.
KW Tetrodotoxin-sensitive sodium channel; rat; PN4 sodium channel; stroke;
KW nervous system disorder; epilepsy; brain injury; diabetic neuropathy;
KW AIDS-associated neuropathy; therapy.
OS Rattus sp.
PN W09838302-A2.
PD 03-SEP-1998.
PF 20-FEB-1998; E00997.
PR 26-FEB-1997; US-039447.
PA (HOFF) HOFFMANN LA ROCHE & CO AG F.
PI Delgado SG, Dietrich PS, Fish LM, Herman RC, Sangameswaran L;
DR WPI: 98-481204/41.
DR N-PSDB: V58420.
PT New rat tetrodotoxin-sensitive sodium channel alpha subunit and DNA
PT - for detecting inhibitors which alleviate pain, and treating
PT nervous system disorders, e.g. epilepsy, stroke, diabetic and AIDS
PT neuropathy
PS Claim 33; Page 44-51; 87pp; English.
CC This sequence is encoded by the rat PN4 sodium channel cDNA clone of the
CC invention. The DNA sequence was isolated from a peripheral nerve from
CC a rat dorsal ganglia. The PN4 sodium channel sequences are
CC tetrodotoxin-sensitive sodium channels. The protein is used in assays for
CC detecting inhibitors of tetrodotoxin-sensitive sodium channels, which
CC alleviate pain. The probes can be used to detect and isolate the DNA or
CC protein in tissues. The antibodies can also be used to isolate the
CC protein. The protein is used as a therapeutic target for compounds to
CC treat disorders of the nervous system, such as epilepsy, stroke and brain
CC injury, diabetic neuropathy, and AIDS-associated neuropathy, etc.
SQ Sequence 1988 AA;

Query Match 86.7%; Score 39; DB 1; Length 1988;
Best Local Similarity 33.3%; Pred. No. 3.92e+02;
Matches 4; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

Db 193 PNWMLDFSVIMM 204
QY 2 PXXXXXXAVILM 13
```



Db 170 PLNSQVTIILM 181  
 QY 2 PXXXXXXAVILM 13

## RESULT 8

ID P80698 standard; protein; 297 AA.  
 AC P80698;  
 DE Recombinant protein which cross-reacts with Simian rotavirus SA-11  
 DE major outer capsid protein (VP7)  
 KW Simian rotavirus SA-11 (ATCC VR999); rotavirus SA-11 gene 9; pAR91;  
 KW rotavirus SA-11 major outer capsid protein VP7; diagnostic reagent;  
 KW baculovirus-insect cell expression system; vaccine;  
 KW Simian rotavirus SA-11 major outer capsid protein (VP7).  
 OS Simian rotavirus SA-11 (ATCC VR999).  
 PN EP-251467-A.  
 PD 07-JAN-1988.  
 PF 19-MAY-1987; 304436.  
 PR 20-JUN-1986; US-876518.  
 PA (ABBO) Abbott Laboratories.  
 PI Smith RE, McConigal T;  
 DR WPI; 88-001354/01.  
 DR N-PSDB; n81280.  
 PT Recombinant rota-virus outer capsid sub-unit protein -  
 PT produced in a baculovirus-insect cell expression system and used  
 PT in diagnostic, vaccine and antibody prodn.  
 PS Example; Fig 3; 49pp; English.  
 CC Rotavirus SA-11 RNA gene segments were used as templates to synthesise  
 CC DNA copies which were subsequently cloned into plasmid vectors in E.coli  
 CC A cloned DNA copy of SA-11 gene 9 was identified and verified by DNA  
 CC sequence analysis to contain the entire protein coding sequence of the  
 CC major outer capsid protein (VP7). SA-11 gene 9 DNA was then subcloned for  
 CC baculovirus expression. The cloned DNA was expressed under control  
 CC of the polyhedrin promoter to yield a subunit protein antigenically  
 CC similar to the native protein, by infection of an insect cell line with  
 CC recombinant baculovirus contg gene 9. The resulting protein contains the  
 CC AA sequence shown in p80698. The resulting expressed gene prod. has been  
 CC shown to be antigenic by reacting with neutralising polyclonal antisera  
 CC derived against rotavirus particles. It can be used in a vaccine or as a  
 CC diagnostic reagent. Recombinant subunit protein of the major outer capsid  
 CC protein (VP7) of rotavirus SA-11 produced in a baculovirus-insect cell  
 CC expression system, the recombinant subunit protein having a molecular  
 CC weight of about 34 kilodaltons and being reactive with neutralising  
 CC antisera against rotavirus SA-11 is claimed.  
 SQ Sequence 297 AA;

Query Match 86.7%; Score 39; DB 1; Length 297;  
 Best Local Similarity 33.3%; Pred. No. 3.92e+02;  
 Matches 4; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

Db 102 POLYCDYNVILM 113  
 QY 2 PXXXXXXAVILM 13

## RESULT 9

ID P01221 standard; protein; 325 AA.  
 AC P91221;  
 DE Rhesus rotavirus (first entry).  
 DE Rhesus rotavirus gene 9 VP7 protein.  
 KW VP7 protein; Rhesus rotavirus; antibodies; rotaviruses; vaccine;  
 KW gene 9.  
 OS Rhesus rotavirus.  
 FH Key Location/Qualifiers  
 FT misc\_difference 69  
 FT region 1. .23  
 FT region 32. .63  
 PN W0806971-A.  
 PD 10-AUG-1989.  
 PF 03-JAN-1989; U00018.  
 PR 01-FEB-1988; US-150670.

PA (STRD) Leland Stanford Jr. Univ.  
 PI Greenberg HB, Mackow E;  
 DR WPI; 89-248894/34.  
 PT Rhesus rotavirus VP3 or VP7 polypeptide(s)  
 PT - which stimulate prodn. of neutralising antibodies for  
 PT protection against rotavirus infection.  
 PS Disclosure; Table 5; 50pp; English.  
 CC Rhesus rotavirus gene 9 VP7 protein (see N91635 - also for variants).  
 CC This, or its active fragments, stimulate prodn. of  
 CC antibodies that neutralise a broad spectrum of rotaviruses, and  
 CC can be used as vaccines. They may be joined to, eg  
 CC beta-galactosidase or Salmonella flagellin. The misc. feature  
 CC is a potential N-linked glycosylation site, and the regions indicate  
 CC hydrophobic areas. See also VP3 (N91636).  
 SQ Sequence 325 AA;

Query Match 86.7%; Score 39; DB 1; Length 325;  
 Best Local Similarity 33.3%; Pred. No. 3.92e+02;  
 Matches 4; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

Db 131 POLYCDYNVILM 142  
 QY 2 PXXXXXXAVILM 13

RESULT 10  
 ID P71565 standard; protein; 326 AA.  
 AC P71565;  
 DT 24-MAY-1991 (first entry)  
 DE Sequence of rotavirus major outer shell protein VP7 serotype SA11.  
 KW Diarrhoea; gastrointestinal disorder; RNA virus; vaccine.  
 OS Rotavirus.  
 FH Key Location/Qualifiers  
 FT region 164. .295  
 FT /label= 14K polypeptide  
 FT /note= A fragment of VP7 with this SQ is claimed"

PN AU8666987-A.  
 PD 02-JUL-1987.  
 PF 23-DEC-1986; 017981.  
 PR 26-DEC-1985; US-813661.  
 PR 03-SEP-1986; US-903325.  
 PA (UYSA-) UNIV OF SASKATCHEWA.  
 PI Sabara MJ, Frenchick PJ, Potter AA, Ijaz MK, Gilchrist JE;  
 DR WPI; 87-228567/33.  
 PT New peptide fragments of rota-viral proteins - useful conjugates  
 PT in vaccines for protecting against gastrointestinal disorders and  
 PT diarrhoea.  
 PS Disclosure; Fig 1; 84pp; English.  
 CC The peptide fragments of glycoprotein VP7 and proteins VP6 and VP3  
 CC of rotaviruses are useful when attached to carriers as vaccines for  
 CC birds and mammals, including man. The vaccines confer protection  
 CC against gastrointestinal disorders and diarrhoea produced by the  
 CC rotaviruses. For use in vaccines the peptides are covalently linked  
 CC to eg, keyhole limpet haemocyanin, BSA, ovalbumin, poly-L-lysine,  
 CC or VP6 bovine rotavirus protein. An adjuvant may be included.  
 SQ Sequence 326 AA;

Query Match 86.7%; Score 39; DB 1; Length 326;  
 Best Local Similarity 33.3%; Pred. No. 3.92e+02;  
 Matches 4; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

Db 131 POLYCDYNVILM 142  
 QY 2 PXXXXXXAVILM 13

## RESULT 11

ID P60547 standard; protein; 326 AA.  
 AC P60547;  
 DT 24-JUN-1991 (first entry)  
 DE Segment 8 clone of the hu/5 rotavirus.  
 KW Hu/Australia/5/77 rotavirus; vaccine.  
 OS Human/5 rotavirus.

```

RESULT 5
ID R27737 standard; Protein; 416 AA.
AC R27737;
DE 09-MAR-1993 (first entry)
DE Sequence transcribed from third reading frame of
DE vaccinia virus DNA from positions 17201-18450.
KW virus vector; vaccinia virus; papillomavirus; HPV;
KW immunotherapeutic; neutral site.
OS Vaccinia virus.
PN W09216636-A.
PD 01-OCT-1992. G00424.
PF 10-MAR-1992; G00424.
PR 14-MAR-1991; GB-005383.
PA (IMMU ) IMMUNOLOGY LTD.
PI Boursnell MEG, Inglis SC, Munro AJ;
DR WPI: 92-349219/42.
DR N-PSDB: Q29467.
PT Recombinant virus vectors encoding human papillomavirus proteins
PT - for treating and vaccinating against HPV infections and
PT conditions caused by them, such as cervical cancer
PS Disclosure: Fig 19: 83pp; English.
CC To make a recombinant virus vector comprising human papillomavirus
CC genes inserted into the vaccinia virus genome, neutral sites
CC for insertion must be utilised such that replicative ability is not
CC adversely affected. The neutral sites are identified by analysing
CC the viral genome to identify ORFs which are likely to encode
CC functional genes and selecting sites between such ORFs or within
CC sequences for non-functional genes. The sequence shown is that
CC transcribed from the vaccinia virus WR strain positions 17201-18450
CC contg. the regions covered by the four fragments SaI, G, H and I.
CC The sequence was transcribed in all three reading frames to determine
CC genuine vaccinia virus genes via codon usage, thus determining neutral
CC sites. HPV DNA sequences may be inserted neutral sites, e.g. those
CC encoding E6 or E7 of HPV 16 and 18 or mutants of these proteins.
CC The recombinant virus vector may be used immunotherapeutically to
CC activate cells of the immune system against HPV.
CC See also R27723-43.
SQ Sequence 416 AA;

Query Match 88.9%; Score 40; DB 1; Length 416;
Best Local Similarity 33.3%; Pred. No. 3.02e+02;
Matches 4; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

Db 211 PHTVTTPIILM 222
QY 2 PXXXXXXAVILM 13

RESULT 6
ID R36780 standard; Protein; 1365 AA.
AC R36780;
DE 16-JUL-1993 (first entry)
DE KRE5.
DE Yeast; cell wall; beta-glucan; assembly; pathway; KRE1; KRE5; growth;
DE secretory; O-linked mannose; (1>6)-beta-glucan; epistasis; morphology;
DE hydrophilic; glycoprotein; COOH-terminal; endoplasmic reticulum; ER;
DE retention signal; antifungal agent.
OS Saccharomyces cerevisiae.
PN US5194600-A.
PD 16-MAR-1993.
PF 05-MAR-1990; 488316.
PR 05-MAR-1990; US-488316.
PA (ROYA-) ROYAL INST ADVANCEMENT LEARNING.
PI Boone C, Bussey H, Hill K, Meaden P, Sommer SS;
DR WPI: 93-109384/13.
DR N-PSDB: Q38899.
PT New DNA encoding genes which participate in beta-glucan assembly
PT - useful for producing mutants for in-vivo screening of
PT antifungal agents and providing tools for in-vitro screening
PS Claim 1; Columns 38-44, 24pp; English.
CC The sequences given in R34785 and R36780 represent proteins which
CC participate in a yeast cell wall beta-glucan assembly pathway.

```

```

CC These proteins represent KRE1 and KRE5 respectively, and are
CC essential for normal cell growth. KRE1 is a Ser/Thr rich protein
CC that is directed into the yeast secretory pathway, where it is
CC highly modified, probably through addition of O-linked mannose
CC residues. Gene disruption of the KRE1 locus leads to a 40% reduced
CC level of cell wall (1>6)-beta-glucan. Mutations at KRE5 also caused
CC defects in cell wall (1>6)-beta-glucan production and appears to be
CC epistatic to KRE1. KRE5 is a large hydrophilic secretory glyco-
CC protein which contains the COOH-terminal endoplasmic reticulum (ER)
CC retention signal (His-Asp-Glu-Leu). Deletion of the KRE5 gene results
CC in cells with aberrant morphology and extremely compromised growth.
CC KRE1 and KRE5 are useful as tools for the in vitro screening of anti-
CC fungal agents which inhibit fungi pathogenic to plants and animals.
CC The genes can be used to produce mutants for in vivo screening of
CC antifungal agents.
SQ Sequence 1365 AA;

Query Match 88.9%; Score 40; DB 1; Length 1365;
Best Local Similarity 41.7%; Pred. No. 3.02e+02;
Matches 5; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

```

```

Db 907 PLKFNVIQVILM 918
QY 2 PXXXXXXAVILM 13

```

```

RESULT 7
ID W76251 standard; Protein; 182 AA.
AC W76251;
DE 02-DEC-1998 (first entry)
DE Human TACE-like protein.
DE TACE; TNF-alpha converting enzyme; tumour necrosis factor; screening;
KW metalloproteinase; antagonist; diagnosis; overexpression; infection;
KW inflammation; immune system; neurological disease.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Peptide 1..15
FT /label= signal
FT Protein 16..182
FT /label= TACE_like_protein
FT Region 169..182
FT /label= transmembrane_region
PN W09831818-A2.
PD 23-JUL-1998.
PF 20-JAN-1998; U00783.
PR 01-AUG-1997; US-054541.
PR 21-JAN-1997; US-034205.
PR 13-JUN-1997; US-049607.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Brewer L, Gentz R, Ni J, Rosen CA, Ruben SM;
DR WPI: 98-414114/35.
DR N-PSDB: V61632.
PT Isolated nucleic acid encoding human metallo-protease(s) - used for
PT diagnosis, treatment and prevention of, e.g. cancer, inflammation,
PT neurological disease and infections
PS Claim 1a; Fig 1; 81pp; English.
CC This sequence represents a novel human TACE-like protein which is a
CC tumour necrosis factor (TNF) alpha converting enzyme and member of the
CC metalloproteinase family. This protein can be used in assays to screen
CC for agonists and antagonists and the nucleic acid is used as a probe
CC for gene mapping, in situ hybridisation and detection of corresponding
CC genes in human tissue, and as sources of probes and primers for
CC diagnosis. The protein and its antigenic fragments are used to raise
CC antibodies (Ab) (which can be used for diagnosis in usual immunoassays
CC or for in vivo imaging) and to screen for (ant)agonists. Antagonists
CC of this protein are used to treat disorders associated with
CC overexpression of TNF-alpha, e.g. inflammation, immune system or
CC neurological diseases, or infection.
SQ Sequence 182 AA;

Query Match 86.7%; Score 39; DB 1; Length 182;
Best Local Similarity 33.3%; Pred. No. 3.92e+02;
Matches 4; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

```

CC drugs which reduce virulence or compounds useful for preventing,  
 CC ameliorating or treating infections in animals or plants.  
 SQ Sequence 4473 AA;

Query Match 95.6%; Score 43; DB 1; Length 4473;  
 Best Local Similarity 41.7%; Pred. No. 1.37e+02;  
 Matches 5; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

Db 1480 QSLILFAVILM 1491

QY 2 PXXXXXXAVILM 13

RESULT 2

ID W81107 standard; Protein; 205 AA.

AC W81107;

DT 16-FEB-1999 (first entry)

DE Human Rho.

KW Human; Rho protein; cell proliferation; inflammation;

KW transplantation; cancer; gene therapy.

OS Homo sapiens.

PN WO9846754-A1.

PD 22-OCT-1998.

PF 16-APR-1998; U07865.

PR 17-APR-1997; US-842976.

PA (INCY-) INCYTE PHARM INC.

PI Goli SK, Hillman JL;

DR WPI; 98-609916/51.

DR N-PSDB; V68232.

PT New isolated human Rho protein - used to develop products for

PT treating, e.g. infections or cancers or inflammation resulting from

PT AIDS, allergies and asthma

PS Claim 1; Fig 1; 59pp; English.

CC The purified human Rho protein (HRHO) is 205 amino acids. The

CC expression of HRHO is associated with cell proliferation and

CC inflammation. HRHO agonists can be used to stimulate cell proliferation,

CC e.g. in heterologous or autologous transplantation or for fighting

CC infection or a cancer or to correct a genetic defect in a disease such as

CC sickle cell anemia, beta thalassemia, cystic fibrosis and Huntington's

CC chorea, or for promoting regeneration or differentiation of cells.

CC Antagonists of HRHO can be used for treating or preventing inflammation

CC or cancer. The PNS can also be used for gene therapy. The products can

CC also be used for detection, diagnosis and drug screening.

SQ Sequence 205 AA;

Query Match 91.1%; Score 41; DB 1; Length 205;

Best Local Similarity 41.7%; Pred. No. 2.33e+02;

Matches 5; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Db 83 PLSYPDTDVILM 94

QY 2 PXXXXXXAVILM 13

RESULT 3

ID W85035 standard; Protein; 442 AA.

AC W85035;

DT 08-FEB-1999 (first entry)

DE Green fluorescent protein-RhoA fusion product.

DE Human; RhoA gene; fusion protein; green fluorescent protein; GFP;

KW Intracellular signalling; chimera.

OS Chimeric - Aequorea victoria.

OS Chimeric - Homo sapiens.

PN WO9845704-A2.

PD 15-OCT-1998.

PF 07-APR-1998; DK0145.

PR 07-APR-1997; DK-000392.

PA (NOVO) NOVO-NORDISK AS.

PI Kasper A, Petersen Bjorn S, Scudder K, Thastrup O,

PI Tullin S;

DR WPI; 98-594491/50.

DR N-PSDB; C71080.

PT Determining effect on signalling pathways in live cells from

PT redistribution of luminophores - specifically fusions of green  
 PT fluorescent protein with a signalling component, and new apparatus,  
 PT particularly for identifying toxins and potential therapeutic agents  
 PS Example 21; Pages 248-249; 326pp; English.  
 CC The present sequence represents a green fluorescent protein (GFP)-human  
 CC RhoA fusion protein. The fusion protein is used in an assay to  
 CC exemplify the invention. The specification describes how quantitative  
 CC information about the influence of a molecule on a cellular response is  
 CC obtained by recording the variation, caused by the molecule, on  
 CC mechanically intact living cells, in the spatially distributed light  
 CC emitted from a luminophore present in the cells. The variation in light  
 CC emission is processed to provide information that correlates spatial  
 CC distribution to the degree of the molecule. The method is used to  
 CC identify agents that (in)directly affect intracellular signalling,  
 CC especially to screen for potential therapeutic agents or toxins, and to  
 CC identify new drug targets.  
 SQ Sequence 442 AA;

Query Match 91.1%; Score 41; DB 1; Length 442;

Best Local Similarity 41.7%; Pred. No. 2.33e+02;

Matches 5; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Db 320 PLSYPDTDVILM 331

QY 2 PXXXXXXAVILM 13

RESULT 4

ID R27731 standard; peptide; 44 AA.

AC R27731;

DT 09-MAR-1993 (first entry)

DE Salf20R.

KW Virus vector; vaccinia virus; papillomavirus; HPV;

KW immunotherapeutic.

OS Vaccinia virus.

PN WO9216636-A.

PD 01-OCT-1992.

PF 10-MAR-1992; G00424.

PR 14-MAR-1991; GB-005383.

PA (IMMU) IMMUNOLOGY LTD.

PI Boursnell MEG, Inglis SC, Munro AJ;

DR WPI; 92-349219/42.

DR N-PSDB; Q29392.

PT Recombinant virus vectors encoding human papillomavirus proteins

PT - for treating and vaccinating against HPV infections and

PT conditions caused by them, such as cervical cancer

PS Disclosure; Fig 10; 83pp; English.

CC To make a recombinant virus vector comprising human papillomavirus

CC genes inserted into the vaccinia virus genome, neutral sites

CC for insertion must be utilised such that replicative ability is not

CC adversely affected. The neutral sites are identified by analysing

CC the viral genome to identify ORFs which are likely to encode

CC functional genes and selecting sites between such ORFs or within

CC sequences for non-functional genes. One such neutral site is site B,

CC present in an intergenic region between Salf20R and Salf20.5R. It is

CC placed 70 bases upstream of Salf20.5R to avoid promoter elements

CC associated with that gene. Also there is no transfection

CC termination signal with which site B could interfere and hence the

CC sequence is suitable as a neutral insertion site. HPV DNA sequences

CC may be inserted at this site, e.g. those encoding E6 or E7 of HPV 16

CC and 18 or mutants of these proteins. The recombinant virus vector may

CC be used immunotherapeutically to activate cells of the immune system

CC against HPV. See also R27723-43.

SQ Sequence 44 AA;

Query Match 88.9%; Score 40; DB 1; Length 44;

Best Local Similarity 33.3%; Pred. No. 3.02e+02;

Matches 4; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

Db 18 PHTYTPILM 29

QY 2 PXXXXXXAVILM 13

\*\*\*\*\*  
W P S R L A  
\*\*\*\*\*  
(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Sat Apr 15 01:18:57 2000; MasPar time 3.28 Seconds  
Tabular output not generated. 93.850 Million cell updates/sec

Title: >US-08-452-843-22  
Description: (1-13) from US08452843.pep  
Perfect Score: 45  
Sequence: 1 XPXXXXXAVILM 13

Scoring table: PAM 150  
Gap 15

Searched: 188963 seqs, 23686106 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: a-geneseqp  
1:geneseqp

Statistics: Mean 14.003; Variance 40.070; scale 0.349

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	43	95.6	4473	1 R97244	Virulence gene cluster	1.37e+02
2	41	91.1	205	1 W81107	Human Rho	2.33e+02
3	41	91.1	442	1 W85035	Green fluorescent prot	2.33e+02
4	40	88.9	44	1 R27731	SalF20R.	3.02e+02
5	40	88.9	416	1 R27737	Sequence transcribed f	3.02e+02
6	40	88.9	1365	1 R36780	KRE5.	3.02e+02
7	39	86.7	182	1 W6251	Human TACE-like protei	3.92e+02
8	39	85.7	297	1 P80698	Recombinant protein wh	3.92e+02
9	39	86.7	325	1 P91221	Rhesus rotavirus gene	3.92e+02
10	39	86.7	326	1 P71565	Sequence of rotavirus	3.92e+02
11	39	86.7	326	1 P60547	Segment 8 clone of the	3.92e+02
12	39	86.7	1479	1 W44119	Human type C lectin.	3.92e+02
13	39	86.7	1978	1 W69361	Tetradotoxin-sensitive	3.92e+02
14	39	86.7	1988	1 W69362	Tetradotoxin-sensitive	3.92e+02
15	38	84.4	37	1 R82534	IE beta chain fragment	5.07e+02
16	38	84.4	65	1 R82535	Hybrid IE beta chain.	5.07e+02
17	38	84.4	82	1 W95773	DEF chimeric molecule	5.07e+02
18	38	84.4	254	1 R82532	Hybrid IE beta chain.	5.07e+02
19	38	84.4	352	1 P70497	Aspergillus niger adha	5.07e+02
20	38	84.4	417	1 R22095	Phosphoglycerate kinas	5.07e+02
21	38	84.4	418	1 R93247	PGK.	5.07e+02
22	38	84.4	418	1 R22025	A. chrysogenum phospho	5.07e+02
23	38	84.4	463	1 W19800	Glycyl-tRNA synthetase	5.07e+02

24	37	82.2	132	1 W16323	Human ARF-p19, a novel	6.53e+02
25	37	82.2	153	1 W73358	S. colwelliana M1ga pr	6.53e+02
26	37	82.2	153	1 R87527	Mel-linked m1ga gene p	6.53e+02
27	37	82.2	165	1 W38723	Streptococcus pneumonia	6.53e+02
28	37	82.2	245	1 W64220	Human secreted protein	6.53e+02
29	37	82.2	318	1 W95502	C. acetobutylicum lnco	6.53e+02
30	37	82.2	325	1 W52296	CRFB4 protein.	6.53e+02
31	37	82.2	326	1 R99720	Swine rotavirus gp38.	6.53e+02
32	37	82.2	326	1 R38696	PSY565 swine rotavirus	6.53e+02
33	37	82.2	326	1 P70499	Pig rota virus gp38 ge	6.53e+02
34	37	82.2	326	1 P94797	Swine rotavirus gp38 G	6.53e+02
35	37	82.2	363	1 R47557	ILRV thymidine kinase.	6.53e+02
36	37	82.2	394	1 Y00876	Human LAPH-1 protein s	6.53e+02
37	37	82.2	485	1 W15280	AUX1 polypeptide invol	6.53e+02
38	37	82.2	525	1 W99453	B.diminuta pimelyl CoA	6.53e+02
39	37	82.2	526	1 W97814	Human butyrophilin.	6.53e+02
40	37	82.2	2016	1 W23994	Human hhl sodium chann	6.53e+02
41	37	82.2	2019	1 R67913	Cardiac sodium channel	6.53e+02
42	37	82.2	2020	1 R06584	Cardiac sodium channel	6.53e+02
43	36	80.0	429	1 R94561	Human adenylyl cyclase	8.41e+02
44	36	80.0	538	1 W99345	Human endogenous retro	8.41e+02
45	36	80.0	1091	1 R28822	Alpha 6B integrin subu	8.41e+02

ALIGNMENTS

RESULT 1  
ID R97244 standard; Protein; 4473 AA.  
AC R97244;  
DT 07-JAN-1997 (first entry)  
DE Virulence gene cluster polypeptide product.  
KW Mutant; adaptation; virulence factor; identification; screening;  
KW vaccine; drugs; infection; treatment.  
OS Salmonella typhimurium.  
FH Key Location/Qualifiers  
FT region  
FT /note= "All x's in this sequence correspond to  
termination codons in the virulence gene  
cluster sequence given in T09224."

WO9617951-A2.  
13-JUN-1996.  
11-DEC-1995; G02875.  
09-DEC-1994; GB-024921.  
31-JAN-1995; GB-001881.  
05-MAY-1995; GB-009239.  
(RPMs-) RPMs TECHNOLOGY LTD.  
PI Holden DM;  
DR WPI; 96-287194/29.  
DR N-PSDB; T09224.  
PT Identifying virulence genes in microorganisms - by introducing  
mutants with insertion inactivated genes into environment and  
retrieval and analysis of mutants  
PS Claim 51. Figure 11; 131pp; English.  
CC A method for identifying a microorganism having a reduced adaptation  
to a particular environment comprising the steps of: (1) providing a  
plurality of microorganisms each of which is independently mutated by  
the insertional inactivation of a gene which a nucleic acid comprising  
a unique marker sequence so that each mutant contains a different  
marker sequence, or clones of the said microorganism; (2) providing  
individually a stored sample of each mutant produced by step (1) and  
providing individually stored nucleic acid comprising the unique  
marker sequence from each individual mutant; (3) introducing a  
plurality of mutants produced by step (1) into the said particular  
environment and allowing those microorganisms which are able to do so  
to grow in the said environment; (4) retrieving microorganisms from  
the said environment or a selected part thereof and isolating the  
CC nucleic acid from the retrieved microorganisms; (5) comparing any  
CC marker sequences in the nucleic acid isolated in step (4) to the  
CC unique marker sequence of each individual mutant stored as in step  
CC (2); and (6) selecting an individual mutant which does not contain any  
CC of the marker sequences as isolated in step (4). The products and  
CC methods can be used for identifying virulence genes in microorganisms.  
CC The mutant microorganisms can be used in vaccines or to screen for

41 #status predicted\  
#modified\_site ADP-ribosylasparagine (Asn) (by botulinum  
exoenzyme C3) #status experimental\  
190 #modified\_site methyl ester carboxyl end (Cys) (in  
mature form) #status predicted\  
190 #binding\_site geranyl-geranyl (Cys) (covalent) #status  
predicted

SUMMARY #length 193 #molecular-weight 21768 #checksum 5051

Query Match 91.1%; Score 41; DB 1; Length 193;  
Best Local Similarity 41.7%; Pred. No. 3.06e+01;  
Matches 5; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Db 71 PLSYPDTDVILM 82  
Qy 2 PXXXXXXAVILM 13

Search completed: Sat Apr 15 01:20:09 2000  
Job time : 16 secs.

```

R.M.; Settlement, J.
EMBO J. (1995) 14:292-302
#journal      Characterization of rho GTPase family homologues in
#title       Drosophila melanogaster: overexpressing Rho1 in retinal
#accession   S54294
#status      preliminary
#molecule_type mRNA
##residues   1-192 ##label HAR
##cross-references EMBL:L38311; NID:g624239; PID:g624240
GENETICS
#gene        FlyBase:Rho1
##cross-references FlyBase:FBgn0014020
CLASSIFICATION #superfamily ras transforming protein; translation elongation
factor Tu homology
GTP binding; P-loop
KEYWORDS      #region nucleotide-binding motif A (P-loop)\
12-19         #region GTP-binding NKXD motif\
117-120       #region GTP-binding SAK/L motif\
160-162       #binding_site Mg-GTP (Lys, Thr, Thr, Asn, Lys, Asp, Ser)
18,19,37,117,118,
120,160       #status predicted
SUMMARY       #length 192 #molecular-weight 21723 #checksum 3721
Query Match   91.1%; Score 41; DB 2; Length 192;
Best Local Similarity 41.7%; Pred. NO. 3.06e+01;
Matches 5; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
Db 71 PLSYPDPTDVL 82
QY 2 PXXXXXXAVIL 13
RESULT 14
ENTRY      G38625      #type complete
TITLE      GTP-binding protein o-rho - electric ray (Discopyge ommata)
ORGANISM   #formal_name Discopyge ommata
DATE       23-Aug-1991 #sequence_revision 23-Aug-1991 #text_change
19-Dec-1998
ACCESSIONS G38625
REFERENCE   Ngsee, J.K.; Elferink, L.A.; Scheller, R.H.
#authors    J. Biol. Chem. (1991) 266:2675-2680
#journal    A family of ras-like GTP-binding proteins expressed in
#title      electromotor neurons.
#cross-references MUID:91115900
#accession   G38625
#status      preliminary
#molecule_type mRNA
##residues   1-192 ##label NGS
##cross-references GB:M38396; NID:g213104; PID:g213105
CLASSIFICATION #superfamily ras transforming protein; translation elongation
factor Tu homology
P-loop
KEYWORDS      #region nucleotide-binding motif A (P-loop)
12-19         #length 192 #molecular-weight 21480 #checksum 2470
SUMMARY       Query Match 91.1%; Score 41; DB 2; Length 192;
Best Local Similarity 41.7%; Pred. NO. 3.06e+01;
Matches 5; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
Db 71 PLSYPDPTDVL 82
QY 2 PXXXXXXAVIL 13
RESULT 15
ENTRY      TVB012      #type complete
TITLE      GTP-binding protein rhoA - bovine
ALTERNATE_NAMES #gb: ras-related homolog A; transforming protein rhoA
ORGANISM       #formal_name Bos primigenius taurus #common_name cattle
15-Jan-1993 #sequence_revision 24-May-1996 #text_change
19-Dec-1998
ACCESSIONS A33518; A60050; A32119; A38324
REFERENCE   Ogorochi, T.; Nemoto, Y.; Nakajima, M.; Nakamura, E.;
#authors    Fujiwara, M.; Narumiya, S.
#journal    Biochem. Biophys. Res. Commun. (1989) 163:1175-1181
#title      cDNA cloning of Gb, the substrate for botulinum
ADP-ribosyltransferase from bovine adrenal gland and its
identification as a rho gene product.
#cross-references MUID:89391974
#accession   A33518
#molecule_type mRNA
##residues   1-193 ##label OGO
##cross-references GB:M27278; NID:g162742; PID:g162743
REFERENCE   A60050
#authors    Hoshijima, M.; Kondo, J.; Kikuchi, A.; Yamamoto, K.; Takai,
Y.
#journal    Brain Res. Mol. Brain Res. (1990) 7:9-16
#title      Purification and characterization from bovine brain membranes
of a GTP-binding protein with a M-r of 21,000.
ADP-ribosylated by an ADP-ribosyltransferase contaminated
in botulinum toxin type CI - identification as the rhoA
gene product.
#cross-references MUID:90135940
#accession   A60050
#molecule_type protein
##residues   8-15;28-43;52-58;100-104;169-181 ##label HOS
REFERENCE   A32119
#authors    Narumiya, S.; Sekine, A.; Fujiwara, M.
#journal    J. Biol. Chem. (1988) 263:17255-17257
#title      Substrate for botulinum ADP-ribosyltransferase, Gb, has an
amino acid sequence homologous to a putative rho gene
product.
#cross-references MUID:89034241
#accession   A32119
#molecule_type protein
##residues   19,'X',21-25;'X',42-50;52-57;59-82,'X',84-97;99-104;'M',
133-158,'X',160-162;169-176 ##label NAR
#note       the amino end of the mature protein is blocked
REFERENCE   A38324
#authors    Williamson, K.C.; Smith, L.A.; Moss, J.; Vaughan, M.
#journal    J. Biol. Chem. (1990) 265:20807-20812
#title      Guanine nucleotide-dependent ADP-ribosylation of soluble rho
catalyzed by Clostridium botulinum C3
ADP-ribosyltransferase. Isolation and characterization of a
newly recognized form of rhoA.
#cross-references MUID:91063876
#accession   A38324
#molecule_type protein
##residues   28-70;99-104 ##label WIL
REFERENCE   A33190
#authors    Sekine, A.; Fujiwara, M.; Narumiya, S.
#journal    J. Biol. Chem. (1989) 264:8602-8605
#title      Asparagine residue in the rho gene product is the
modification site for botulinum ADP-ribosyltransferase.
#cross-references MUID:89255316
#contents    annotation; identification of ADP-ribosylation site
CLASSIFICATION #superfamily ras transforming protein; translation elongation
factor Tu homology
KEYWORDS      acetylated amino end; GTP binding; lipoprotein; membrane
protein; methylated carboxyl end; P-loop; prenylated
cysteine; proto-oncogene; transforming protein
FEATURE       #product GTP-binding protein rhoA #status predicted
2-190        #label MAT\
12-19        #region nucleotide-binding motif A (P-loop)\
117-120      #region GTP-binding NKXD motif\
160-162      #region GTP-binding SAK/L motif\
2            #modified_site acetylated amino end (Ala) (in mature
form) #status predicted\
18,19,37,117,118,
120,160      #binding_site Mg-GTP (Lys, Thr, Thr, Asn, Lys, Asp, Ser)

```

```

Db 36 PLSYPDVTILM 47
|
|
|
QY 2 PXXXXXXAVILM 13

RESULT 10
ENTRY #type complete
TITLE translation elongation factor EF-1, subunit beta homolog -
ORGANISM Archaeoglobus fulgidus
DATE 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change
13-Sep-1998
ACCESSIONS F69321
REFERENCE AG9250
#authors Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson,
K.E.; Ketchum, K.A.; Dodson, R.J.; Gwinn, M.; Hickey, E.K.;
Peterson, J.D.; Richardson, D.L.; Kerlavage, A.R.; Graham,
D.E.; Kyrpides, N.C.; Fleischmann, R.D.; Quackenbush, J.;
Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.;
Dougherty, B.A.; McKenny, K.; Adams, M.D.; Loftus, B.;
Peterson, S.; Reich, C.I.; McNeil, L.K.; Badger, J.H.;
Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman,
J.F.; McDonald, L.; Utterback, T.; Cotton, M.D.; Spriggs,
T.; Artiach, P.; Kaine, B.P.; Sykes, S.M.; Sadow, P.W.;
D'Andrea, K.P.; Bowman, C.; Fujii, C.; Garland, S.A.;
Mason, T.M.; Olsen, G.J.; Fraser, C.M.; Smith, H.O.; Woese,
C.R.; Venter, J.C.
#journal Nature (1997) 390:364-370
#title The complete genome sequence of the hyperthermophilic,
sulfate-reducing archaeon Archaeoglobus fulgidus.
#cross-references M01D:198049343
#accession F69321
##status preliminary; nucleic acid sequence not shown;
translation not shown
##molecule_type DNA
##residues 1-88 #label KLE
##cross-references GB:AE001065; GB:AE000782; NID:g2689388; PID:g2650050;
TIGR:AF0574
CLASSIFICATION #superfamily Sulfolobus solfataricus translation elongation
factor aEF-1 beta
#length 88 #molecular-weight 9809 #checksum 878
SUMMARY
Query Match 91.1%; Score 41; DB 2; Length 88;
Best Local Similarity 33.3%; Pred. No. 3.06e+01;
Matches 4; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

Db 43 PIAFGLKAVILM 54
|
|
|
QY 2 PXXXXXXAVILM 13

RESULT 11
ENTRY TVGAAC #type complete
TITLE transforming protein rho - California sea hare
ORGANISM #formal_name Aplysia californica #common_name California sea
hare
DATE 27-Nov-1985 #sequence_revision 27-Nov-1985 #text_change
19-Dec-1998
ACCESSIONS A01373
REFERENCE A01372
#authors Madaule, P.; Axel, R.
#journal Cell (1985) 41:31-40
#title A novel ras-related gene family.
#cross-references M01D:85201682
#accession A01373
##molecule_type mRNA
##residues 1-192 #label MAD
##cross-references GB:M10078; NID:gi155803; PID:gi155804
COMMENT This protein is homologous with the ras transforming proteins.
GENETICS
#gene rho
CLASSIFICATION #superfamily ras transforming protein; translation elongation
factor Tu homology

```

---

```

KEYWORDS GTP binding; lipoprotein; membrane protein; methylated
carboxyl end; P-loop; prenylated cysteine; proto-oncogene;
transforming protein
FEATURE
12-19 #region nucleotide-binding motif A (P-loop)\
117-120 #region GTP-binding NKXD motif\
160-162 #region GTP-binding SAK/L motif\
18,19,37,117,118, #binding_site Mg-GTP (Lys, Thr, Thr, Asn, Lys, Asp, Ser)
120,160 #status predicted\
189 #binding_site geranyl-geranyl (Cys) (covalent) #status
predicted\
189 #modified_site methyl ester carboxyl end (Cys) (in
mature form) #status predicted
SUMMARY #length 192 #molecular-weight 21661 #checksum 5045
Query Match 91.1%; Score 41; DB 1; Length 192;
Best Local Similarity 41.7%; Pred. No. 3.06e+01;
Matches 5; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
Db 71 PLSYPDVTILM 82
|
|
|
QY 2 PXXXXXXAVILM 13

RESULT 12
ENTRY A55492 #type complete
TITLE GTP-binding protein rhoA - Caenorhabditis elegans
ORGANISM #formal_name Caenorhabditis elegans
DATE 03-Mar-1995 #sequence_revision 03-Mar-1995 #text_change
19-Dec-1998
ACCESSIONS A55492
REFERENCE A55492
#authors Chen, W.; Lim, L.
#journal J. Biol. Chem. (1994) 269:32394-32404
#title The Caenorhabditis elegans small GTP-binding protein RhoA is
enriched in the nerve ring and sensory neurons during
larval development.
#accession A55492
##status preliminary
##molecule_type mRNA
##residues 1-192 #label CHE
#cross-references GB:I36965; NID:g558500; PID:g558501
CLASSIFICATION #superfamily ras transforming protein; translation elongation
factor Tu homology
KEYWORDS GTP binding; P-loop
FEATURE
12-19 #region nucleotide-binding motif A (P-loop)\
117-120 #region GTP-binding NKXD motif\
160-162 #region GTP-binding SAK/L motif\
18,19,37,117,118, #binding_site Mg-GTP (Lys, Thr, Thr, Asn, Lys, Asp, Ser)
120,160 #status predicted
SUMMARY #length 192 #molecular-weight 21635 #checksum 2714
Query Match 91.1%; Score 41; DB 2; Length 192;
Best Local Similarity 41.7%; Pred. No. 3.06e+01;
Matches 5; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
Db 71 PLSYPDVTILM 82
|
|
|
QY 2 PXXXXXXAVILM 13

RESULT 13
ENTRY S54294 #type complete
TITLE Rho1 protein - fruit fly (Drosophila melanogaster)
ORGANISM #formal_name Drosophila melanogaster
DATE 15-Jul-1995 #sequence_revision 01-Sep-1995 #text_change
19-Dec-1998
ACCESSIONS S54294
REFERENCE S54294
#authors Hariharan, I.K.; Hu, K.Q.; Asha, H.; Quintanilla, A.; Ezzeil,

```

## ##experimental\_source strain K-12

```
GENETICS
#gene      acrF: envD
CLASSIFICATION
#superfamily acriflavin resistance protein
KEYWORDS
FEATURE
12-28      #domain transmembrane #status predicted #label TM1\
343-359    #domain transmembrane #status predicted #label TM2\
370-386    #domain transmembrane #status predicted #label TM3\
397-413    #domain transmembrane #status predicted #label TM4\
442-458    #domain transmembrane #status predicted #label TM5\
472-488    #domain transmembrane #status predicted #label TM6\
540-556    #domain transmembrane #status predicted #label TM7\
874-890    #domain transmembrane #status predicted #label TM8\
898-914    #domain transmembrane #status predicted #label TM9\
974-990    #domain transmembrane #status predicted #label TM10\
1012-1028  #domain transmembrane #status predicted #label TM11\
SUMMARY    #length 1034 #molecular-weight 111454 #checksum 9723

Query Match      95.6%; Score 43; DB 2; Length 1034;
Best Local Similarity 41.7%; Pred. No. 1.11e+01;
Matches 5; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

Db 9 PIFAWLAIILM 20
|
|
|
|
|
QY 2 PXXXXXXAVILM 13

RESULT 6
ENTRY #type fragment
TITLE probable transforming protein ras (clone PSH-RHO2) - penaeid
ORGANISM shrimp (Penaeus monodon) (fragment)
DATE 27-Jan-1995 #sequence_revision 26-Jul-1996 #text_change
19-Dec-1998

ACCESSIONS S42051
REFERENCE S42050
#authors Gendreau, S.; Lee, R.; Mialhe, E.
#submission submitted to the EMBL Data Library, February 1994
#accession S42051
#molecule_type DNA
#residues 1-34 #label GEN
##cross-references EMBL:Z30081; NID:G454998; PID:G454999
CLASSIFICATION #superfamily ras transforming protein; translation elongation
factor Tu homology
transforming protein
KEYWORDS #length 34 #checksum 6009
SUMMARY

Query Match      91.1%; Score 41; DB 2; Length 34;
Best Local Similarity 41.7%; Pred. No. 3.06e+01;
Matches 5; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Db 5 PLSYPDTDVILM 16
|
|
|
|
|
QY 2 PXXXXXXAVILM 13

RESULT 7
ENTRY #type fragment
TITLE probable transforming protein ras (clone PSH-RHO1) - penaeid
shrimp (Penaeus monodon) (fragment)
ORGANISM #formal_name Penaeus monodon
DATE 27-Jan-1995 #sequence_revision 26-Jul-1996 #text_change
19-Dec-1998

ACCESSIONS S42050
REFERENCE S42050
#authors Gendreau, S.; Lee, R.; Mialhe, E.
#submission submitted to the EMBL Data Library, February 1994
#accession S42050
#molecule_type DNA
#residues 1-34 #label GEN
##cross-references EMBL:Z30080; NID:G454996; PID:e98778; PID:gl335659
CLASSIFICATION #superfamily ras transforming protein; translation elongation
factor Tu homology
transforming protein
```

```
factor Tu homology
transforming protein
#length 34 #checksum 5632

Query Match      91.1%; Score 41; DB 2; Length 34;
Best Local Similarity 41.7%; Pred. No. 3.06e+01;
Matches 5; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Db 5 PLSYPDTDVILM 16
|
|
|
|
|
QY 2 PXXXXXXAVILM 13

RESULT 8
ENTRY #type fragment
TITLE rhoA protein - pig (fragment)
ORGANISM #formal_name Sus scrofa domestica #common_name domestic pig
DATE 31-Jan-1997 #sequence_revision 31-Jan-1997 #text_change
19-Dec-1998

ACCESSIONS PC4222
REFERENCE PC4222
#authors Nishimura, J.; Sakihara, C.; Zhou, Y.; Kanaide, H.
#journal Biochem. Biophys. Res. Commun. (1996) 227:750-754
#title Expression of rho A and rho kinase mRNAs in porcine vascular
smooth muscle.
#cross-references MUID:97040692
#accession PC4222
#molecule_type mRNA
#residues 1-66 #label NIS
##cross-references DBJ:D8949; NID:91695730; PID:dl014663; PID:91695731
COMMENT This protein is involved in the inhibition of myosin light chain
phosphatase.
CLASSIFICATION #superfamily ras transforming protein; translation elongation
factor Tu homology
transforming protein
SUMMARY #length 66 #checksum 9673

Query Match      91.1%; Score 41; DB 2; Length 66;
Best Local Similarity 41.7%; Pred. No. 3.06e+01;
Matches 5; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Db 36 PLSYPDTDVILM 47
|
|
|
|
|
QY 2 PXXXXXXAVILM 13

RESULT 9
ENTRY #type fragment
TITLE rho A protein - rat (fragment)
ORGANISM #formal_name Rattus norvegicus #common_name Norway rat
DATE 28-May-1997 #sequence_revision 18-Jul-1997 #text_change
19-Dec-1998

ACCESSIONS PC4266
REFERENCE PC4266
#authors Niino, N.; Nishimura, J.; Sakihara, C.; Nakano, H.; Kanaide,
H.
#journal Biochem. Biophys. Res. Commun. (1997) 230:356-359
#title Up-regulation of rho A and rho-kinase mRNAs in the rat
myometrium during pregnancy.
#cross-references MUID:97168976
#accession PC4266
#molecule_type mRNA
#residues 1-66 #label NII
COMMENT This protein is involved in the Ca2+ sensitivity of the smooth
muscle myofilaments. It inhibits myosin light chain
phosphorylation.
CLASSIFICATION #superfamily ras transforming protein; translation elongation
factor Tu homology
transforming protein
SUMMARY #length 66 #checksum 9673

Query Match      91.1%; Score 41; DB 2; Length 66;
Best Local Similarity 41.7%; Pred. No. 3.06e+01;
Matches 5; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
```



```

#journal      Gene (1993) 130:247-251
#title       Characterization of an Aspergillus nidulans genomic DNA
             fragment conferring phosphate-non-repressible
             acid-phosphatase activity.
#cross-references MUID:93366181
#accession   JN0784
             #molecule_type DNA
             #residues      1-113 #label MAC
             #cross-references GB:M96993; NID:g168001; PID:g168002
GENETICS
#introns     65/2
SUMMARY      #length 113 #molecular-weight 12638 #checksum 2363
Query Match 95.6%; Score 43; DB 2; Length 113;
Best Local Similarity 41.7%; Pred. No. 1.11e+01;
Matches 5; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

Db 57 PIVAHETVILM 68
QY 2 PXXXXXXAVILM 13

RESULT 3
ENTRY   B71017 #type complete
TITLE   hypothetical protein PH1431 - Pyrococcus horikoshii
ORGANISM
#formal_name Pyrococcus horikoshii
DATE    14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change
ACCESSIONS B71017
REFERENCE   A71000
#authors   Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.;
             Hino, Y.; Yamamoto, S.; Sekine, M.; Baba, S.; Kosugi, H.;
             Hosoyama, A.; Nagai, Y.; Sakai, M.; Ogura, K.; Otsuka, R.;
             Nakazawa, H.; Takamiya, M.; Ohfuku, Y.; Funahashi, T.;
             Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kishida, N.; Oguchi,
             A.; Aoki, K.; Yoshizawa, T.; Nakamura, Y.; Robb, F.T.;
             Horikoshi, K.; Masuchi, Y.; Shizuya, H.; Kikuchi, H.
#journal   DNA Res. (1998) 5:95-76
#title     Complete sequence and gene organization of the genome of a
             hyper-thermophilic archaeobacterium, Pyrococcus horikoshii
             OT3.
#cross-references MUID:98344137
#accession B71017
#status    preliminary; nucleic acid sequence not shown;
             translation not shown
             #molecule_type DNA
             #residues      1-510 #label KAW
             #cross-references GB:AP000006; NID:g3236133; PID:d1031481; PID:g3257855
             #experimental_source strain OT3
             #note          this accession replaces an interim accession for a
                     sequence replaced by GenBank

GENETICS
#gene      PH1431
SUMMARY    #length 510 #molecular-weight 55287 #checksum 3776
Query Match 95.6%; Score 43; DB 2; Length 510;
Best Local Similarity 41.7%; Pred. No. 1.11e+01;
Matches 5; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

Db 466 PMFILVIAILM 477
QY 2 PXXXXXXAVILM 13

RESULT 4
ENTRY   C70879 #type complete
TITLE   Probable ftsK - Mycobacterium tuberculosis (strain H37Rv)
ORGANISM
#formal_name Mycobacterium tuberculosis
DATE    17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change
ACCESSIONS C70879
REFERENCE   A70500
#authors   Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher,

```

```

C.; Harris, D.; Gordon, S.V.; Eiglmeier, K.; Gas, S.; Barry
III, C.E.; Tekalia, F.; Badcock, K.; Basham, D.; Brown, D.;
Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.;
Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;
Hornsby, T.; Jagels, K.; Krogh, A.; McLean, J.; Moule, S.;
Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.;
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.;
Skellton, S.; Squares, S.; Squires, R.; Sulston, J.E.;
Taylor, K.; Whitehead, S.; Barrell, B.G.
#journal    Nature (1998) 393:537-544
#title      Deciphering the biology of Mycobacterium tuberculosis from
             the complete genome sequence.
#cross-references MUID:98295987
#accession C70879
#status    preliminary; nucleic acid sequence not shown;
             translation not shown
             #molecule_type DNA
             #residues      1-883 #label COL
             #cross-references GB:AL008967; GB:AL123456; NID:g3261491; PID:e1173878;
                     PID:g2624270
             ##experimental_source strain H37Rv
GENETICS
#gene      ftsK
SUMMARY    #length 883 #molecular-weight 94405 #checksum 311
Query Match 95.6%; Score 43; DB 2; Length 883;
Best Local Similarity 41.7%; Pred. No. 1.11e+01;
Matches 5; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

Db 193 PLVAAAVAVILM 204
QY 2 PXXXXXXAVILM 13

RESULT 5
ENTRY   D65119 #type complete
TITLE   acriflavin resistance protein acrF - Escherichia coli
ALTERNATE_NAMES
ORGANISM #formal_name Escherichia coli
DATE    12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change
ACCESSIONS D65119; S18537
REFERENCE   A64720
#authors   Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.;
             Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.;
             Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.;
             Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao,
             Y.
#journal   Science (1997) 277:1453-1462
#title     The complete genome sequence of Escherichia coli K-12.
#cross-references MUID:97426617
#accession D65119
#status    nucleic acid sequence not shown; translation not shown
             #molecule_type DNA
             #residues      1-1034 #label BLAT
             #cross-references GB:AE000405; GB:U000096; NID:g1789659; PID:g1789666;
                     UWGP:B3266
             ##experimental_source strain K-12, substrain MG1655
REFERENCE   S18536
#authors   Klein, J.R.; Henrich, B.; Plapp, R.
#journal   Mol. Gen. Genet. (1991) 230:230-240
#title     Molecular analysis and nucleotide sequence of the envC
             operon of Escherichia coli.
#cross-references MUID:92079901
#accession S18537
#status    preliminary
             #molecule_type DNA
             #residues      69-222, 'R', 224-283, 'T', 285, 'PP', 288-337,
                     'KCKNAVRSYAGYGVDSVLAETASADPHHCGRGVDRVDCVCHPRFLLHQH
                     TNDVRDGCACR', 404-422, 'RYV', 426-485, 'F', 487-787,
                     'PLCPQ', 793-836, 'A', 838-846, 'RQNYC', 853, 'H', 855-1034
                     #label KLE
#cross-references EMBL:X57948; NID:g510827; PID:g510830

```

\*\*\*\*\*  
W P L H  
(TM)  
\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Sat Apr 15 01:19:53 2000; MasPar time 3.27 Seconds  
Tabular output not generated. 159.292 Million cell updates/sec

Title: >US-08-452-843-22  
Description: (1-13) from US08452843.pep  
Perfect Score: 45  
Sequence: 1 XPXXXXXXAVILM 13

Scoring table: PAM 150  
Gap 15

Searched: 122810 seqs, 40068593 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: p1r62  
1:p1r1 2:p1r2 3:p1r3 4:p1r4

Statistics: Mean 20.294; Variance 22.476; scale 0.903

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	45	100.0	436	D65007	hypothetical protein	3.88e+00
2	43	95.6	113	JN0784	acid phosphatase, 12.9K	1.11e+01
3	43	95.6	510	B71017	hypothetical protein	1.11e+01
4	43	95.6	883	C70879	probable ftsK - Mycob	1.11e+01
5	43	95.6	1034	D65119	acriflavin resistance	1.11e+01
6	41	91.1	34	S42051	probable transforming	3.06e+01
7	41	91.1	34	S42050	probable transforming	3.06e+01
8	41	91.1	66	PC4222	rhoA protein - pig (f	3.06e+01
9	41	91.1	66	PC4256	rho A protein - rat (	3.06e+01
10	41	91.1	88	F69321	translation elongatio	3.06e+01
11	41	91.1	192	1 TVGAAC	transforming protein	3.06e+01
12	41	91.1	192	A54492	GTP-binding protein r	3.06e+01
13	41	91.1	192	S54294	Rho1 protein - fruit	3.06e+01
14	41	91.1	192	G38625	GTP-binding protein o	3.06e+01
15	41	91.1	193	1 TVB012	GTP-binding protein r	3.06e+01
16	41	91.1	193	1 TVH0RC	GTP-binding protein r	3.06e+01
17	41	91.1	193	1 TVH012	GTP-binding protein r	3.06e+01
18	41	91.1	193	2 H36364	GTP-binding protein r	3.06e+01
19	41	91.1	196	1 TVRTRH	GTP-binding protein r	3.06e+01
20	41	91.1	196	2 JC5075	GTP-binding protein r	3.06e+01
21	41	91.1	196	1 TVH0RH	GTP-binding protein r	3.06e+01
22	41	91.1	354	2 D31751	protein kinase cataly	3.06e+01
23	41	91.1	361	1 SYECCR	chorismate synthase (	3.06e+01

24 41 91.1 361 1 SYECCR chorismate synthase ( 3.06e+01  
25 41 91.1 372 2 A55510 chorismate synthase ( 3.06e+01  
26 41 91.1 376 2 E31751 protein kinase cataly 3.06e+01  
27 41 91.1 1090 2 A41696 regulatory protein ni 3.06e+01  
28 40 88.9 103 2 JQ1791 Salp16R protein - vac 5.02e+01  
29 40 88.9 103 2 A42523 A53R protein - vaccin 5.02e+01  
30 40 88.9 782 2 S62583 hypothetical protein 5.02e+01  
31 40 88.9 971 1 JQ1634 outer capsid protein 5.02e+01  
32 40 88.9 1365 1 BVBYK5 killer toxin resistan 5.02e+01  
33 39 86.7 151 2 I38367 small G protein - hum 8.14e+01  
34 39 86.7 326 1 VGXRHH glycoprotein VP7 prec 8.14e+01  
35 39 86.7 326 1 VGXRDS glycoprotein VP7 prec 8.14e+01  
36 39 86.7 326 1 VGXR2S glycoprotein VP7 prec 8.14e+01  
37 39 86.7 326 1 VGXRHN glycoprotein VP7 prec 8.14e+01  
38 39 86.7 326 1 VGXR1S glycoprotein VP7 prec 8.14e+01  
39 39 86.7 326 1 A44891 glycoprotein VP7 prec 8.14e+01  
40 39 86.7 326 1 VGXRHU glycoprotein VP7 prec 8.14e+01  
41 39 86.7 372 2 JQ2135 NADH dehydrogenase (u 8.14e+01  
42 39 86.7 653 2 JQ1241 viral replicase 1 - b 8.14e+01  
43 39 86.7 1040 2 A34695 axonal glycoprotein r 8.14e+01  
44 39 86.7 1835 2 I54323 sodium channel alpha 8.14e+01  
45 39 86.7 1836 2 JS0648 sodium channel alpha 8.14e+01

ALIGNMENTS

RESULT 1  
ENTRY D65007 #type complete  
TITLE hypothetical protein D2342 - Escherichia coli (strain K-12)  
ORGANISM #formal\_name Escherichia coli  
DATE 12-Sep-1997 #sequence\_revision 17-Sep-1997 #text\_change 18-Sep-1998

ACCESSIONS D65007  
REFERENCE A64720  
#authors Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.; Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.; Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao, Y.

#journal Science (1997) 277:1453-1462  
#title The complete genome sequence of Escherichia coli K-12.  
#cross-references MUID:197426617  
#accession D65007  
##status preliminary; nucleic acid sequence not shown;  
##molecule\_type DNA translation not shown  
##residues 1-436 #label BLAT  
##cross-references GB:AE000322; GB:U000096; NID:g1788672; PID:g1788683; UWGP:B2342

CLASSIFICATION #superfamily long-chain-fatty-acid beta-oxidation multienzyme complex beta chain

SUMMARY #length 436 #molecular-weight 46530 #checksum 5449

Query Match 100.0%; Score 45; DB 2; Length 436;  
Best Local Similarity 50.0%; Pred. No. 3.88e+00;  
Matches 6; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 274 PLTDGAAVILM 285  
|  
|  
|  
|  
QY 2 PXXXXXXAVILM 13

RESULT 2  
ENTRY JN0784 #type complete  
TITLE acid phosphatase, 12.9K - Emericella nidulans  
ORGANISM #formal\_name Emericella nidulans, Aspergillus nidulans  
DATE 24-Feb-1994 #sequence\_revision 24-Feb-1994 #text\_change 17-Mar-1999

ACCESSIONS JN0784: S27409  
REFERENCE JN0783  
#authors MacRae, W.D.; Buxton, F.P.; Sibley, S.; Garven, S.; Gwynne, D.I.; Arst Jr., H.N.; Davies, R.W.

THIS PAGE BLANK (USPTO)

```
RT Mycobacterium tuberculosis H37Rv, and comparison with Mycobacterium
RT leprae."
RL PROC. Natl. Acad. Sci. U.S.A. 93:3132-3137(1996).
DR EMBL; AL021897; CAA17177.1; -.
DR PFAM; PF00310; GATase_2; 1.
KW Hypothetical protein.
SQ SEQUENCE 287 AA; 32037 MW; 52A7A15E CRC32;

Query Match 77.6%; Score 45; DB 2; Length 287;
Best Local Similarity 85.7%; Pred. No. 1.65e+00;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 214 RAHSTHL 220
QY 3 RAHSSHL 9

RESULT 13
ID Q53796 PRELIMINARY; PRT; 301 AA.
AC Q53796;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)
DE BLEOMYCIN ACETYLTRANSFERASE.
OS Streptomyces verticillius.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomycetes.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 15003;
RX MEDLINE; 95129853.
RA CALCUTT M.J., SCHMIDT F.J.;
RT "Gene organization in the bleomycin-resistance region of the producer
RT organism Streptomyces verticillius.";
RL Gene 151:17-21(1994).
DR EMBL; L26955; AAB00461.1; -.
DR PFAM; PF00583; Acetyltransf; 1.
KW Transferase.
SQ SEQUENCE 301 AA; 32225 MW; C8A31F16 CRC32;

Query Match 75.9%; Score 44; DB 2; Length 301;
Best Local Similarity 62.5%; Pred. No. 3.02e+00;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 5 PRAHTAHL 12
QY 2 SRAHSSHL 9

RESULT 14
ID Q29495 PRELIMINARY; PRT; 207 AA.
AC Q29495;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)
DE ARALKYLAMINE N-ACETYLTRANSFERASE (EC 2.3.1.87)
DE (ARALKYLAMINE N-ACETYLTRANSFERASE) (SEROTONIN ACETYLTRANSFERASE)
DE (SEROTONIN ACETYLASE).
OS Ovis aries (Sheep).
OC Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Mammalia;
OC Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae;
OC Caprinae; Ovis.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-DORSETT X RAMBOUILLET;
RX MEDLINE; 96099405.
RA COON S.L., ROSEBOOM P.H., BALER R., WELLER J.L., NAMBOODIRI M.A.A.,
RA KOONIN E.V., KLEIN D.C.;
RT "Pineal serotonin N-acetyltransferase: expression cloning and
RT molecular analysis";
RL Science 270:1681-1683(1995).
CC -!- CATALYTIC ACTIVITY: ACETYL-COA + ARALKYLAMINE = COA + N-
ACETYLARALKYLAMINE.
```

```
DR EMBL; U29663; AAC48690.1; -.
DR PFAM; PF00583; Acetyltransf; 1.
KW Transferase; Acyltransferase.
SQ SEQUENCE 207 AA; 23076 MW; 2B4429A1 CRC32;

Query Match 74.1%; Score 43; DB 5; Length 207;
Best Local Similarity 62.5%; Pred. No. 5.48e+00;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 114 PRGSAHL 121
QY 2 SRAHSSHL 9

RESULT 15
ID Q9XZ13 PRELIMINARY; PRT; 496 AA.
AC Q9XZ13;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
DE PUTATIVE NICOTINIC ACETYLCHOLINE RECEPTOR ALPHA 7-1 SUBUNIT.
OS Heliothis virescens (Noctuid moth) (Owlet moth).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
OC Noctuoidea; Noctuidae; Heliothinae; Heliothis.
RN [1]
RP SEQUENCE FROM N.A.
RA SCHULTE T., OELLERS N., ADAMCZEWSKI M.;
RT "Putative alpha subunits of insect nicotinic acetylcholine receptors
RT more similar to vertebrate alpha 7 subunits and C. elegans Ce21 than
RT to other insect nicotinic acetylcholine receptor alpha subunits.";
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE LIGAND-GATED IONIC CHANNELS FAMILY.
DR EMBL; AF143846; AAD32697.1; -.
DR PROSITE; PS00236; NEUOTR_ION_CHANNEL; 1.
KW Receptor; Postsynaptic membrane; Ionic channel; Glycoprotein;
KW Transmembrane.
SQ SEQUENCE 496 AA; 56347 MW; 1EF11E40 CRC32;

Query Match 74.1%; Score 43; DB 5; Length 496;
Best Local Similarity 66.7%; Pred. No. 5.48e+00;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 2 GGRARRSHL 10
QY 1 GSRAHSSHL 9

Search completed: Sat Apr 15 01:15:47 2000
Job time : 93 secs.
```

```
DR EMBL; X60014; CAA42629.1; -
DR HSP; P04637; 1SAH.
DR PFAM; PF00870; P53; 1.
FT VARIANT 237 237 I -> M.
FT NON_TER 393 393
SQ SEQUENCE 393 AA; 43694 MW; 9BB81992 CRC32;

Query Match 100.0%; Score 58; DB 4; Length 393;
Best Local Similarity 100.0%; Pred. No. 3.13e-04;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 361 GSRHSHSL 369
QY 1 GSRHSHSL 9

RESULT 10 PRELIMINARY; PRT; 393 AA.
AC Q16811;
DT 01-NOV-1996 (TRENBLrel. 01, Created)
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
DE CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 85126934.
RA BENCHIMOL S.;
RA MATHIASHEWSKI G.; LAMB P.; PTM D.; PEACOCK J.; CRAWFORD L.;
RT "Isolation and characterization of a human p53 cDNA clone: expression
of the human p53 gene.";
RL EMBO J. 3:3257-3262(1984).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 87064416.
RA LAMB P.; CRAWFORD L.;
RT "Characterization of the human p53 gene.";
RL Mol. Cell. Biol. 6:1379-1385(1986).
CC -1- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT
PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL
CYCLE REGULATION, IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY
REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED
FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF
CYCLIN-DEPENDENT KINASES (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
DR EMBL; M13121; AAA59987.1; JOINED.
DR EMBL; M13112; AAA59987.1; JOINED.
DR EMBL; M13113; AAA59987.1; JOINED.
DR EMBL; M13114; AAA59987.1; JOINED.
DR EMBL; M13115; AAA59987.1; JOINED.
DR EMBL; M13116; AAA59987.1; JOINED.
DR EMBL; M13117; AAA59987.1; JOINED.
DR EMBL; M13118; AAA59987.1; JOINED.
DR EMBL; M13119; AAA59987.1; JOINED.
DR EMBL; M13120; AAA59987.1; JOINED.
DR HSP; P04637; 1TSR.
DR PROSITE; PS00348; P53; 1.
DR PFAM; PF00870; P53; 1.
KW Repeat; Tumor antigen; Anti-oncogene; DNA-binding;
KW Transcription regulation; Activator;
FT NON_TER 393
SQ SEQUENCE 393 AA; 43698 MW; 3EA71431 CRC32;

Query Match 100.0%; Score 58; DB 4; Length 393;
Best Local Similarity 100.0%; Pred. No. 3.13e-04;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 361 GSRHSHSL 369
QY 1 GSRHSHSL 9

RESULT 11 PRELIMINARY; PRT; 393 AA.
AC Q16848;
DT 01-NOV-1996 (TRENBLrel. 01, Created)
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
DE CELLULAR TUMOR ANTIGEN P53.
GN TP53.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 87089826.
RA HARRIS N.; BRILL E.; SHOHAT O.; PROKOCIMER M.; WOLF D.; ARAI N.;
RA ROTTER V.;
RT "Molecular basis for heterogeneity of the human p53 protein.";
RL Mol. Cell. Biol. 6:4650-4656(1986).
CC -1- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT
PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL
CYCLE REGULATION, IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY
REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED
FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF
CYCLIN-DEPENDENT KINASES (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
DR EMBL; M14694; AAA61211.1; -.
DR HSP; P04637; 1TSR.
DR PROSITE; PS00348; P53; 1.
DR PFAM; PF00870; P53; 1.
DR PRINTS; PR00386; P53SUPPRESSR.
DR Nuclear protein; Phosphorylation; Anti-oncogene; DNA-binding;
KW Transcription regulation; Activator.
SQ SEQUENCE 393 AA; 43723 MW; DA7D302F CRC32;

Query Match 100.0%; Score 58; DB 4; Length 393;
Best Local Similarity 100.0%; Pred. No. 3.13e-04;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 361 GSRHSHSL 369
QY 1 GSRHSHSL 9

RESULT 12 PRELIMINARY; PRT; 287 AA.
AC O53409;
DT 01-JUN-1998 (TRENBLrel. 06, Created)
DT 01-JUN-1998 (TRENBLrel. 06, Last sequence update)
DE HYPOTHETICAL 32.0 KD PROTEIN.
GN MTVO17.14.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-H37RV;
RA DEVLIN K.; CHURCHER C.M.;
RA Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN-H37RV;
RA COLE S.T.; PARKHILL J.; BARRELL B.G.; RAJANDREAM M.A.;
RA Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RX STRAIN-H37RV;
RA MEDLINE; 96181548.
RA PHILIPP W.J.; POULET S.; EIGLMEIER K.; PASCOPELLA L.;
RA BALASUBRAMANTAN V.; HEYM B.; BERGH S.; BLOOM B.R.; JACOBS W.R. JR.;
RA COLE S.T.;
RA "An integrated map of the genome of the tubercle bacillus,
```

```
Query Match 100.0%; Score 58; DB 4; Length 393;
Best Local Similarity 100.0%; Pred. No. 3.13e-04;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 361 GSRASHSHL 369
QY 1 GSRASHSHL 9

RESULT 6
ID Q16808 PRELIMINARY; PRT; 393 AA.
AC Q16808;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-NOV-1999 (Tremblrel. 12, Last annotation update)
DE CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).
GN P53.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 92007731.
RA FARRELL P.J., ALLAN G., SHANAHAN F., VOUSDEN K.H., CROOK T.;
RT "p53 is frequently mutated in Burkitt's lymphoma cell lines.";
RL EMBO J. 10:2879-2887(1991).
CC -1- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT
CC PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL
CC CYCLE REGULATION, IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY
CC REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED
CC FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF
CC CYCLIN-DEPENDENT KINASES (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
DR EMBL; X60018; CAA42633.1; -.
DR HSSP; P04637; 1SAH.
DR PFAM; P000870; P53; 1.
DR PROSITE; PS00348; P53; 1.
KW Anti-oncogene; DNA-binding; Transcription regulation; Activator;
KW Nuclear protein; Phosphorylation.
FT VARIANT 163 163 H -> Y.
FT NON_TER 393 393
SQ SEQUENCE 393 AA; 43627 MW; AFD8A9E3 CRC32;

Query Match 100.0%; Score 58; DB 4; Length 393;
Best Local Similarity 100.0%; Pred. No. 3.13e-04;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 361 GSRASHSHL 369
QY 1 GSRASHSHL 9

RESULT 7
ID Q16535 PRELIMINARY; PRT; 393 AA.
AC Q16535;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-NOV-1999 (Tremblrel. 12, Last annotation update)
DE P53 TRANSFORMATION SUPPRESSOR (FRAGMENT).
GN P53.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 92007731.
RA FARRELL P.J., ALLAN G., SHANAHAN F., VOUSDEN K.H., CROOK T.;
RT "p53 is frequently mutated in Burkitt's lymphoma cell lines.";
RL EMBO J. 10:2879-2887(1991).
DR EMBL; X60017; CAA42632.1; -.
DR EMBL; X60015; CAA42630.1; -.
DR HSSP; P04637; 1SAH.
DR PFAM; P000870; P53; 1.

Query Match 100.0%; Score 58; DB 4; Length 393;
Best Local Similarity 100.0%; Pred. No. 3.13e-04;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 361 GSRASHSHL 369
QY 1 GSRASHSHL 9

RESULT 8
ID Q16809 PRELIMINARY; PRT; 393 AA.
AC Q16809;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-NOV-1999 (Tremblrel. 12, Last annotation update)
DE CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).
GN P53.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 92007731.
RA FARRELL P.J., ALLAN G., SHANAHAN F., VOUSDEN K.H., CROOK T.;
RT "p53 is frequently mutated in Burkitt's lymphoma cell lines.";
RL EMBO J. 10:2879-2887(1991).
CC -1- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT
CC PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL
CC CYCLE REGULATION, IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY
CC REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED
CC FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF
CC CYCLIN-DEPENDENT KINASES (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
DR EMBL; X60019; CAA42634.1; -.
DR HSSP; P04637; 1SAH.
DR PROSITE; PS00348; P53; 1.
DR PFAM; P000870; P53; 1.
KW Anti-oncogene; DNA-binding; Transcription regulation; Activator;
KW Nuclear protein; Phosphorylation.
FT VARIANT 213 213 Q -> R.
FT NON_TER 393 393
SQ SEQUENCE 393 AA; 43684 MW; CB70BD7F CRC32;

Query Match 100.0%; Score 58; DB 4; Length 393;
Best Local Similarity 100.0%; Pred. No. 3.13e-04;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 361 GSRASHSHL 369
QY 1 GSRASHSHL 9

RESULT 9
ID Q15087 PRELIMINARY; PRT; 393 AA.
AC Q15087;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-NOV-1999 (Tremblrel. 12, Last annotation update)
DE P53 TRANSFORMATION SUPPRESSOR (FRAGMENT).
GN P53.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 92007731.
RA FARRELL P.J., ALLAN G., SHANAHAN F., VOUSDEN K.H., CROOK T.;
RT "p53 is frequently mutated in Burkitt's lymphoma cell lines.";
RL EMBO J. 10:2879-2887(1991).
```



\*\*\*\*\*  
W P S R L  
\*\*\*\*\*  
(TM)  
\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Sat Apr 15 01:14:14 2000; MasPar time 7.27 Seconds  
Tabular output not generated. 85.783 Million cell updates/sec

Title: >US-08-452-843-21  
Description: (1-9) from US08452843.pep  
Perfect Score: 58  
Sequence: 1 GSRASHSHL 9

Scoring table: PAM 150  
Gap 15

Searched: 225878 seqs, 69334122 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: sptrembl12  
1:sp.archaea 2:sp.bacteria 3:sp.fungi 4:sp.human  
5:sp.invertebrate 6:sp.mammal 7:sp.mhc 8:sp.organelle  
9:sp.phage 10:sp.plant 11:sp.rodent 12:sp.unclassified  
13:sp.vertebrae 14:sp.virus

Statistics: Mean 19.917; Variance 19.359; scale 1.029

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	58	100.0	281	6 Q29475	CELLULAR TUMOR ANTIGEN	3.13e-04
2	58	100.0	393	4 Q15088	P53 TRANSFORMATION SUP	3.13e-04
3	58	100.0	393	4 Q15086	P53 TRANSFORMATION SUP	3.13e-04
4	58	100.0	393	4 Q16810	CELLULAR TUMOR ANTIGEN	3.13e-04
5	58	100.0	393	4 Q16807	CELLULAR TUMOR ANTIGEN	3.13e-04
6	58	100.0	393	4 Q16808	CELLULAR TUMOR ANTIGEN	3.13e-04
7	58	100.0	393	4 Q16535	P53 TRANSFORMATION SUP	3.13e-04
8	58	100.0	393	4 Q16809	CELLULAR TUMOR ANTIGEN	3.13e-04
9	58	100.0	393	4 Q15087	P53 TRANSFORMATION SUP	3.13e-04
10	58	100.0	393	4 Q16811	CELLULAR TUMOR ANTIGEN	3.13e-04
11	58	100.0	393	4 Q16848	CELLULAR TUMOR ANTIGEN	3.13e-04
12	45	77.6	287	2 Q53409	HYPOTHETICAL 32.0 KD P	1.65e+00
13	44	75.9	301	2 Q53796	BLEOMYCIN ACETYLTRANSFERASE	3.02e+00
14	43	74.1	207	6 Q29495	ARYLALYLAMINE N-ACETYLTRANSFERASE	5.48e+00
15	43	74.1	496	5 Q9X213	PUTATIVE NICOTINIC ACETYLCHOLINE RECEPTOR	5.48e+00
16	43	74.1	1181	4 Q9Y526	DJ439F8.2 (NOVEL KIAA)	5.48e+00
17	42	72.4	286	14 P93032	P53 (FRAGMENT)	9.85e+00
18	42	72.4	286	14 P93003	P53 (FRAGMENT)	9.85e+00
19	42	72.4	378	14 P93002	P53 (FRAGMENT)	9.85e+00
20	42	72.4	384	14 Q88523	GLYCOPROTEIN HOMOLOGUE	9.85e+00

21	42	72.4	390	11 O70366	CELLULAR TUMOR ANTIGEN	9.85e+00
22	42	72.4	519	10 Q65573	PUTATIVE PRL1 ASSOCIAT	9.85e+00
23	42	72.4	944	4 Q14163	KIAA0150 PROTEIN (FRAG	9.85e+00
24	42	72.4	1265	4 Q94899	KIAA0807 PROTEIN (FRAG	9.85e+00
25	42	72.4	3268	3 Q03280	D8035.1P (UBIQUITIN LI	9.85e+00
26	41	70.7	93	2 Q47549	PURM GENE ENCODING 5'-	1.75e+01
27	41	70.7	208	2 Q86755	HYPOTHETICAL 23.3 KD P	1.75e+01
28	41	70.7	383	5 Q76138	DIHYDROOROTASE.	1.75e+01
29	41	70.7	400	2 Q49949	UI756G.	1.75e+01
30	41	70.7	405	14 Q11374	HYPOTHETICAL 44.0 KD P	1.75e+01
31	41	70.7	473	10 Q43035	1-AMINOCYCLOPROPANE-1-	1.75e+01
32	41	70.7	533	11 Q35240	PROTON GATED CATION CH	1.75e+01
33	41	70.7	567	5 Q21495	MC3D4.6 PROTEIN.	1.75e+01
34	41	70.7	642	14 Q98310	M0144R.	1.75e+01
35	41	70.7	688	2 Q86090	POLYPHOSPHATE KINASE.	1.75e+01
36	41	70.7	977	4 Q75300	ES/130.	1.75e+01
37	41	70.7	1018	11 Q54782	MANNOSEDASE 2, ALPHA B	1.75e+01
38	41	70.7	1534	6 Q28298	RIBOSOME RECEPTOR.	1.75e+01
39	40	69.0	327	6 Q97491	FAS PROTEIN.	3.09e+01
40	40	69.0	346	8 Q33240	MATURASE (FRAGMENT).	3.09e+01
41	40	69.0	506	8 Q47154	RIBOSOMAL MATURASE.	3.09e+01
42	40	69.0	505	8 Q47143	RIBOSOMAL MATURASE.	3.09e+01
43	40	69.0	584	4 Q60366	HEPATOCYTE GROWTH FACT	3.09e+01
44	40	69.0	1382	11 P97523	HGF RECEPTOR PRECURSOR	3.09e+01
45	40	69.0	1382	11 P97579	HEPATOCYTE GROWTH FACT	3.09e+01

ALIGNMENTS

RESULT	ID	Q29475	PRELIMINARY;	PRT;	281 AA.
AC	Q29475;				
DT	01-NOV-1996	(TrEMBLrel. 01, Created)			
DT	01-NOV-1996	(TrEMBLrel. 01, Last sequence update)			
DT	01-NOV-1999	(TrEMBLrel. 12, Last annotation update)			
DE	CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).				
GN	P53.				
OS	Canis familiaris (Dog).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;				
OC	Eutheria; Carnivora; Fissipedia; Canidae; Canis.				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=MAMMARY GLAND;				
RX	MEDLINE; 97194812.				
RA	VAN LEEUWEN I., RUTTEMAN G.R., HELLMAN E., CORNELISSE C.C.J.,				
RA	DEVILLEE P.;				
RT	"P53 mutations in mammary tumor cell lines and corresponding tumor				
RT	tissues in the dog."				
RL	Anticancer Res. 16:3737-3744(1996).				
CC	-1- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT				
CC	PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL				
CC	CYCLE REGULATION, IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY				
CC	REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED				
CC	FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF				
CC	-1- SUBCELLULAR LOCATION: NUCLEAR.				
CC	EMBL; L37107; AAC37335.1;				
DR	HSSP; P04637; 1SAH.				
DR	PROSITE; PS00348; P53; 1.				
DR	PFAM; PF00870; P53; 1.				
KW	Anti-oncogene; DNA-binding; Transcription regulation; Activator;				
KW	Nuclear protein; Phosphorylation.				
FT	NON_TER 1				
FT	NON_TER 281				
SQ	SEQUENCE 281 AA; 31762 MW; FC7BAE31 CRC32;				

Query Match 100.0%; Score 58; DB 6; Length 281;  
Best Local Similarity 100.0%; Pred. No. 3.13e-04;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Db 254 GSRASHSHL 262  
1 GSRASHSHL 9  
QY



SQ SEQUENCE 998 AA; 110286 MW; 6FB75A43 CRC32;

Query Match 72.4%; Score 42; DB 1; Length 998;  
Best Local Similarity 62.5%; Pred. No. 5.34e+00;  
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 408 PRVHTSHL 415

Oy 2 SRAHSSHL 9  
:|:|

Search completed: Sat Apr 15 01:13:56 2000  
Job time : 42 secs.



[2]  
RN RP SEQUENCE OF 1-14; 44-54; 74-84 AND 106-119, AND ACETYLATION.  
RC STRAIN-S288C;  
RX MEDLINE; 95009940.  
RA BROWN J.D., HANN B.C., MEDZIHRADSKY K.F., NIWA M., BURLINGAME A.L.,  
RA WALTER P.;  
RT Subunits of the Saccharomyces cerevisiae signal recognition particle  
RT required for its functional expression.;  
RL EMBO J. 13:4390-4400(1994).  
CC -!- FUNCTION: SIGNAL-RECOGNITION-PARTICLE ASSEMBLY HAS A CRUCIAL ROLE  
CC IN TARGETING SECRETORY PROTEINS TO THE ROUGH ENDOPLASMIC  
CC RETICULUM MEMBRANE. IT MUST BE INVOLVED INTIMATELY IN THE  
CC TRANSLLOCATION OF A WIDE VARIETY OF PROTEIN SUBSTRATES. SRP21  
CC COULD POSSIBLY BIND TO SCRI.  
CC -!- SUBUNIT: YEAST SIGNAL RECOGNITION PARTICLE CONSISTS OF A 7S RNA  
CC MOLECULE (SCR1) AND AT LEAST SEVEN PROTEIN SUBUNITS: SRP72, SRP68,  
CC SRP54, SEC65, SRP21, SRP14 AND SRP7.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; S44213; AAB23073.1; -;  
DR EMBL; 228122; CA81963.1; -;  
DR PIR; S25360; S25360  
DR SGB; L0002062; SRP21.  
RW Signal recognition particle; Acetylation; RNA-binding.  
FT INIT\_MET 0  
FT MOD\_RES 1 1 ACETYLATION.  
FT DOMAIN 154 166 LYS-RICH (HIGHLY BASIC).  
FT SEQUENCE 166 AA; 18294 MW; 6978BA92 CRC32;  
SQ  
Query Match 72.4%; Score 42; DB 1; Length 166;  
Best Local Similarity 71.4%; Pred. No. 5.34e+00;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
Db 43 RTHNSHL 49  
Qy 3 RAHSSHL 9  
-----  
RESULT 13  
ID YPBG\_BACSU STANDARD; PRT; 259 AA.  
AC P50733;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE HYPOTHETICAL 28.6 KD PROTEIN IN RECO-CMK INTERGENIC REGION PRECURSOR.  
GN YPBG.  
OS Bacillus subtilis.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
OC Bacillus/Staphylococcus group; Bacillus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-168 / MAREBURG;  
RX MEDLINE; 96349105.  
RA SOROKIN A.V., AZEVEDO V., ZUNSTEIN E., GALLERON N., EHRLICH S.D.,  
RA SERROR P.;  
RT "Sequence analysis of the Bacillus subtilis chromosome region between  
RT the serA and kdg loci cloned in a yeast artificial chromosome.";  
RL Microbiology 142:2005-2016(1996).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

CC EMBL; L47648; AAC83951.1; -;  
DR EMBL; 299115; CAB14214.1; -;  
DR EMBL; 299116; CAB14230.1; -;  
DR SUBTILIST; BG11433; YPBG.  
KW Hypothetical protein; ATP-binding; Signal.  
FT SIGNAL 1 19 OR 26 (POTENTIAL).  
FT CHAIN 20 221 HYPOTHETICAL PROTEIN YPBG.  
FT NP\_BIND 214 229 ATP (POTENTIAL).  
SQ SEQUENCE 259 AA; 28560 MW; 0E6EBFF9 CRC32;  
Query Match 72.4%; Score 42; DB 1; Length 259;  
Best Local Similarity 50.0%; Pred. No. 5.34e+00;  
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;  
Db 68 ARSHAPL 75  
Qy 2 SRAHSSHL 9  
-----  
RESULT 14  
ID P53\_MOUSE STANDARD; PRT; 390 AA.  
AC P02340;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 01-NOV-1990 (Rel. 16, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE CELLULAR TUMOR ANTIGEN P53.  
GN TP53 OR TRP53 OR P53.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 85027173.  
RA BIENZ B., ZAKUT-HOURI R., GIVOL D., OREN M.;  
RT "Analysis of the gene coding for the murine cellular tumour antigen  
RT p53.";  
RL EMBO J. 3:2179-2183(1984).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 84068204.  
RA ZAKUT-HOURI R., OREN M., BIENZ B., LAVIE V., HAZUM S., GIVOL D.;  
RT "A single gene and a pseudogene for the cellular tumour antigen p53.";  
RL Nature 306:594-597(1983).  
RN [3]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 84272240.  
RA JENKINS J.R., RUDGE K., REDMOND S., WADE-EVANS A.;  
RT "Cloning and expression analysis of full length mouse cDNA sequences  
RT encoding the transformation associated protein p53.";  
RL Nucleic Acids Res. 12:5609-5626(1984).  
RN [4]  
RP SEQUENCE FROM N.A. (CLONES PCD53; P53-M11 AND P53-M8).  
RX MEDLINE; 87064640.  
RA ARAI N., NOMURA D., YOKOTA K., WOLF D., BRILL E., SHOHAT O.,  
RA ROTTNER V.;  
RT "Immunologically distinct p53 molecules generated by alternative  
RT splicing.";  
RL Mol. Cell. Biol. 6:3232-3239(1986).  
RN [5]  
RP SEQUENCE OF 222-258 FROM N.A.  
RX MEDLINE; 92115342.  
RA BURNS P.A., KEMP C.J., GANNON J.V., LANE D.P., BREMMER R.,  
RA BALMAIN A.;  
RT "Loss of heterozygosity and mutational alterations of the p53 gene in  
RT skin tumours of interspecific hybrid mice.";  
RL Oncogene 6:2363-2369(1991).  
RN [6]  
RP PHOSPHORYLATION SITES.  
RX MEDLINE; 86149247.  
RA SAMAD A., ANDERSON C.W., CARROLL R.B.;  
RT "Mapping of phosphonoester and apparent phosphodiester bonds of the  
RT oncogene product p53 from simian virus 40-transformed 3T3 cells.";

RA MERRITT C.M., BOARD P.G.;

RT "Structure and characterisation of a duplicated human alpha 1 acid

RL glycoprotein gene.";

RN Gene 66:97-106(1988).

RP [3]

RX DISULFIDE BONDS.

RA MEDLINE: 74290014.

RA SCHMID K., BUERGI W., COLLINS J.H., NANNO S.;

RT "The disulfide bonds of alaph-acid glycoprotein.";

RL Biochemistry 13:2694-2697(1974).

RN [4]

RP CARBOHYDRATE-BINDING SITES.

RA MEDLINE: 92231810.

RA TREUHEIT M.J., COSTELLO C.E., HALSALL H.B.;

RT "Analysis of the five glycosylation sites of human alpha 1-acid

RL glycoprotein.";

CC Biochem. J. 293:105-112(1992).

CC -!- FUNCTION: APPEARS TO FUNCTION IN MODULATING THE ACTIVITY OF THE

CC IMMUNE SYSTEM DURING THE ACUTE-PHASE REACTION.

CC -!- INDUCTION: ALPHA-1-AGP IS SYNTHESIZED IN THE LIVER, THE

CC SYNTHESIS BEING CONTROLLED BY GLUCOCORTICOID, INTERLEUKIN-1

CC AND INTERLEUKIN-6. IT INCREASES 5- TO 50-FOLD UPON INFLAMMATION.

CC -!- POLYMORPHISM: MANY DIFFERENT VARIANTS OF ORM2 ARE KNOWN.

CC -!- SIMILARITY: BELONGS TO THE LIPOCALIN FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL: M20615; AAD15302.2;

DR EMBL: M20611; AAD15302.2; JOINED.

DR EMBL: M20612; AAD15302.2; JOINED.

DR EMBL: M20613; AAD15302.2; JOINED.

DR EMBL: M20614; AAD15302.2; JOINED.

DR EMBL: M20610; AAD15303.1; ALT\_SEQ.

DR EMBL: M20675; AAD15303.1; JOINED.

DR EMBL: M20606; AAD15303.1; JOINED.

DR EMBL: M20607; AAD15303.1; JOINED.

DR EMBL: M20608; AAD15303.1; JOINED.

DR EMBL: M20609; AAD15303.1; JOINED.

DR EMBL: M20675; CAA29874.1; JOINED.

DR EMBL: X06674; CAA29874.1; ALT\_SEQ.

DR EMBL: X06676; CAA29875.1; ALT\_SEQ.

DR EMBL: X05780; NOT\_ANNOTATED\_CDS.

DR EMBL: X06678; CAA29877.1; -.

DR EMBL: X06679; CAA29878.1; -.

DR EMBL: X06680; CAA29879.1; -.

DR EMBL: X05784; NOT\_ANNOTATED\_CDS.

DR EMBL: M21540; AAA51549.1; -.

DR PIR: J03326; J03326.

DR PIR: B28346; B28346.

DR SWISS-2DPAGE; P19652; HUMAN.

DR MIM: 138610; -.

DR PROSITE: PS00213; LIPOCALIN; 1.

DR PFAM: PF00061; lipocalin; 1.

KW Glycoprotein; Plasma; Acute phase; Signal; Lipocalin;

KW Multigene family.

FT SIGNAL 1 18

FT CHAIN 19 201 ALPHA-1-ACID GLYCOPROTEIN 2.

FT MOD\_RES 19 19 PYRROLIDONE CARBOXYLIC ACID.

FT DISULFID 23 165

FT DISULFID 90 183

FT CARBOHYD 33 33

FT CARBOHYD 56 56

FT CARBOHYD 72 72

FT CARBOHYD 93 93

FT CARBOHYD 103 103

SEQUENCE 201 AA; 23602 MW; 105ED8CF CRC32;

Query Match 74.1%; Score 43; DB 1; Length 201;

Best Local Similarity 55.6%; Pred. No. 2.98e+00;

Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 111 GGREHVAHL 119

QY 1 GSRAHSHL 9

|||||

RESULT 11

ID NULM\_STRPU STANDARD; PRT; 97 AA.

AC P15554;

DT 01-APR-1990 (Rel. 14, Created)

DT 01-APR-1990 (Rel. 14, Last sequence update)

DT 01-FEB-1996 (Rel. 33, Last annotation update)

DE NADH-UBIQUINONE OXIDOREDUCTASE CHAIN 4L (EC 1.6.5.3).

GN ND4L.

OS Strongylocentrotus purpuratus (Purple sea urchin).

OG Mitochondrion.

OC Eukaryota; Metazoa; Echinodermata; Echinozoa; Echinoidea;

OC Euechinoidea; Echinacea; Echinoida; Strongylocentrotidae;

OC Strongylocentrotus.

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE; 89011951.

RA JACOBS H.T., ELLIOTT D.J., MATH V.B., FARQUHARSON A.;

RT "Nucleotide sequence and gene organization of sea urchin

RT mitochondrial DNA.";

RL J. Mol. Biol. 202:185-217(1988).

CC -!- CATALYTIC ACTIVITY: NADH + UBIQUINONE - NAD(+) + UBIQUINOL.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL: X12631; CAA31154.1; ALT\_SEQ.

DR PIR: S01502; S01502.

DR PFAM: PF00420; Oxidored\_q2; 1.

KW Oxidoreductase; NAD; Ubiquinone; Mitochondrion.

SQ SEQUENCE 97 AA; 10610 MW; 74E4F9B4 CRC32;

Query Match 72.4%; Score 42; DB 1; Length 97;

Best Local Similarity 75.0%; Pred. No. 5.34e+00;

Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 81 SRTHSSNL 88

QY 2 SRAHSHL 9

|||||

RESULT 12

ID SR21\_YEAST STANDARD; PRT; 166 AA.

AC P32342;

DT 01-OCT-1993 (Rel. 27, Created)

DT 01-FEB-1995 (Rel. 31, Last sequence update)

DT 01-FEB-1995 (Rel. 31, Last annotation update)

DE SIGNAL RECOGNITION PARTICLE 21 KD PROTEIN (SRP21).

GN SRP21 OR YKL122C OR YKL527.

OS Saccharomyces cerevisiae (Baker's yeast).

OC Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;

OC Saccharomycetaceae; Saccharomyces.

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE; 92383952.

RA COLLEAUX L., RICHARD G.-F., THIERRY A., DUJON B.;

RT "Sequence of a segment of yeast chromosome XI identifies a new

RT mitochondrial carrier, a new member of the G protein family, and a

RT protein with the PAACK motif of the H1 histones.";

RT Yeast 8:325-336(1992).

```
RX MEDLINE; 95352829.
RA DEQUIEDT F., KETTMANN R., BURNY A., WILLEMS L.;
RT "Nucleotide sequence of the bovine P53 tumor-suppressor cDNA.";
RL DNA seq. 5:261-264(1995).
RN [2]
RP SEQUENCE OF 13-386 FROM N.A.
RC SPECIES=BOVINE; STRAIN=HOLSTEIN; TISSUE=THYMUS;
RX MEDLINE; 96401400.
RA KOMORI H., ISHIGURO N., HORIUCHI M., SHINAGAWA M., AIDA Y.;
RT "Predominant p53 mutations in enzootic bovine leukemic cell lines.";
RL Vet. Immunol. Immunopathol. 52:53-63(1996).
RN [3]
RP SEQUENCE FROM N.A.
RC SPECIES=B-INDICUS; STRAIN=BORAN; TISSUE=BLOOD;
RA BISHOP R.R.P., GOBRIGT E.E.I.;
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES
CC GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL
CC CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN
CC TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION, IT IS A
CC TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION
CC BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF
CC THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.
CC APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF
CC BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2
CC EXPRESSION.
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY
CC OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED
CC IN MANY TYPES OF CANCER.
CC -1- SIMILARITY: BELONGS TO THE P53 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X81704; CAA57348.1;
DR EMBL; D49825; BAA08629.1;
DR EMBL; U74486; AAB51214.1;
DR HSSP; P04637; 1YCR.
DR PFAM; PF00870; P53; 1.
KW Anti-oncogene; DNA-binding; Transcription regulation; Activator;
FT DOMAIN 1 59 ASP/GLU-RICH (ACIDIC).
FT MOD_RES 304 316 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT CONFLICT 385 385 PHOSPHORYLATION (BY SIMILARITY).
FT CONFLICT 380 380 R -> T (IN REF. 2).
SQ SEQUENCE 386 AA; 43255 MW; 0322BF3D CRC32;

Query Match 87.9%; Score 51; DB 1; Length 386;
Best Local Similarity 100.0%; Pred. No. 2.01e-02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 355 SRAHSHSL 362
QY 2 SRAHSHSL 9
|||||||
RESULT 9
ID P53_RABIT STANDARD; PRT; 391 AA.
AC Q95330.
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE CELLULAR TUMOR ANTIGEN P53.
GN TP53.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrate; Mammalia;

Eutheria; Lagomorpha; Leporidae; Oryctolagus.
[1]
RN SEQUENCE FROM N.A.
RC STRAIN=NEW ZEALAND;
RX MEDLINE; 97208869.
RA LE GOAS F., MAY P., RONCO P., CARON DE FROMENTEL C.;
RT "CDNA cloning and immunological characterization of rabbit p53.";
RL Gene 185:169-173(1997).
CC -1- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES
CC GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL
CC CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN
CC TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION, IT IS A
CC TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION
CC BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF
CC THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.
CC APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF
CC BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2
CC EXPRESSION (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY
CC OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED
CC IN MANY TYPES OF CANCER.
CC -1- SIMILARITY: BELONGS TO THE P53 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X90592; CAA62216.1;
DR HSSP; P04637; 1YCR.
DR PROSITE; PS00348; P53; 1.
DR PFAM; PF00870; P53; 1.
KW Anti-oncogene; DNA-binding; Transcription regulation; Activator;
FT DOMAIN 1 70 ASP/GLU-RICH (ACIDIC).
FT MOD_RES 308 321 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT MOD_RES 390 390 PHOSPHORYLATION (BY SIMILARITY).
SQ SEQUENCE 391 AA; 43435 MW; 30A36172 CRC32;

Query Match 84.5%; Score 49; DB 1; Length 391;
Best Local Similarity 88.9%; Pred. No. 7.39e-02;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 359 GSAHSSYL 367
QY 1 GSAHSSYL 9
|||||||
RESULT 10
ID ALAH_HUMAN STANDARD; PRT; 201 AA.
AC P19652; Q16571;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE ALPHA-1-ACID GLYCOPROTEIN 2 PRECURSOR (AGP 2) (OROSOMUCOID 2) (OMD 2).
GN ORM2 OR AGP2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrate; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
[1]
RN SEQUENCE FROM N.A.
RX MEDLINE; 88029318.
RA DENTE L., PIZZA M.G., METSPALU A., CORTESE R.;
RT "Structure and expression of the genes coding for human alpha 1-acid
RT glycoprotein.";
RL EMBO J. 6:2289-2296(1987).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 88329732.
```

RA OLSCHWANG S., LAURENT-PUIG P., VASSAL A., SALMON R.-J., THOMAS G.:  
RT "Characterization of a frequent polymorphism in the coding sequence  
RT of the TP53 gene in colonic cancer patients and a control  
RT population";  
RL Hum. Genet. 86:369-370(1991).  
RN [20]  
RP VARIANT LFS THR-133.  
RX MEDLINE; 92034774.  
RA LAW J.C., STRONG L.C., CHIDAMBARAM A., FERRELL R.E.:  
RT "A germ line mutation in exon 5 of the p53 gene in an extended cancer  
RT family";  
RL Cancer Res. 51:6385-6387(1991).  
RN [21]  
RP VARIANTS LFS CYS-245; TRP-248; PRO-252 AND LYS-258.  
RX MEDLINE; 91057657.  
RA MALKIN D., LI F.P., STRONG L.C., FRAUMENI J.F. JR., NELSON C.E.,  
RA KIM D.H., KASSEL J., GRYKA M.A., BISCHOFF F.Z., TAINSKY M.A.,  
RA FRIEND S.H.:  
RT "Germ line p53 mutations in a familial syndrome of breast cancer,  
RT sarcomas, and other neoplasms";  
RL Science 250:1233-1238(1990).  
RN [22]  
RP VARIANT LFS ASP-245.  
RX MEDLINE; 91080929.  
RA SRIVASTAVA S., ZOU Z., PINOLLO K., BLATTNER W., CHANG E.H.:  
RT "Germ-line transmission of a mutated p53 gene in a cancer-prone  
RT family with Li-Fraumeni syndrome";  
RL Nature 348:747-749(1990).  
RN [23]  
RP VARIANT LFS LEU-272.  
RX MEDLINE; 92147883.  
RA FELIX C.A., NAU M.M., TAKAHASHI T., MITSUDOMI T., CHIBA I.,  
RA POPLACK D.G., REAMAN G.H., COLE D.E., LETTERIO J.J., WHANG-PENG J.,  
RA KNUTSEN T., MINNA J.D.:  
RT "Hereditary and acquired p53 gene mutations in childhood acute  
RT lymphoblastic leukemia";  
RL J. Clin. Invest. 89:640-647(1992).  
RN [24]  
RP VARIANTS LFS HIS-273 AND VAL-325.  
RX MEDLINE; 92228023.  
RA MALKIN D., JOLLY K.W., BARBIER N., LOOK A.T., FRIEND S.H.,  
RA GEBHARDT M.C., ANDERSEN T.I., BORRESEN A.-L., LI F.P., GARBER J.,  
RA STRONG L.C.:  
RT "Germline mutations of the p53 tumor-suppressor gene in children and  
RT young adults with second malignant neoplasms";  
RL New Engl. J. Med. 326:1309-1315(1992).  
RN [25]  
RP VARIANTS BREAST TUMORS GLN-132; SER-249; LYS-280 AND LYS-285.  
RX MEDLINE; 90295284.  
RA BARTEK J., ICGO R., GANNON J., LANE D.P.:  
RT "Genetic and immunochemical analysis of mutant p53 in human breast  
RT cancer cell lines";  
RL Oncogene 5:893-899(1990).  
RN [26]  
RP VARIANTS COLON TUMORS PHE-241 AND HIS-273.  
RX MEDLINE; 91017544.  
RA RODRIGUES N.R., ROWAN A., SMITH M.E.F., KERR I.B., BODMER W.F.,  
RA GANNON J.V., LANE D.P.:  
... Note: remainder of annotations omitted.

Query Match 100.0%; Score 58; DB 1; Length 393;  
Best Local Similarity 100.0%; Pred. No. 1.71e-04;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 361 GSRAHSHL 369  
|||||

QY 1 GSRAHSHL 9

RESULT 7  
ID. P53\_SHEEP STANDARD; PRT; 382 AA.  
AC. P51664;

DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE CELLULAR TUMOR ANTIGEN P53.  
GN TP53.  
OS Ovis aries (Sheep).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae;  
OC Caprinae; Ovis.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=BLOOD;  
RX MEDLINE; 95352828.  
RA DEQUIEDT F., KETTMANN R., BURNY A., WILLEMS L.:  
RT "Nucleotide sequence of the ovine p53 tumor-suppressor cDNA and its  
RT genomic organization";  
RL DNA Seq. 5:255-259(1995).  
CC -!- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES  
CC GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL  
CC CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN  
CC TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION, IT IS A  
CC TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION  
CC BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF  
CC THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.  
CC APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF  
CC BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2  
CC EXPRESSION.  
CC -!- SUBCELLULAR LOCATION: NUCLEAR.  
CC -!- SIMILARITY: BELONGS TO THE P53 FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; X81705; CAA57349.1; -  
DR HSSP; P04637; 1PET.  
DR PROSITE; PS00348; P53; 1.  
DR PFAM; PF00870; P53; 1.  
KW Anti-oncogene; DNA-binding; Transcription regulation; Activator;  
FT Nucleic acid binding; Phosphorylation; Apoptosis.  
FT DOMAIN 1 66 ASP/GLU-RICH (ACIDIC).  
FT MOD\_RES 300 312 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
FT MOD\_RES 381 381 PHOSPHORYLATION (BY SIMILARITY).  
SQ SEQUENCE 382 AA; 42809 MW; 0CB99A00 CRC32;  
  
Query Match 87.9%; Score 51; DB 1; Length 382;  
Best Local Similarity 100.0%; Pred. No. 2.01e-02;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Db 351 SRAHSHL 358  
|||||  
QY 2 SRAHSHL 9  
  
RESULT 8  
ID P53\_BOVIN STANDARD; PRT; 386 AA.  
AC Q29628;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 15-JUL-1998 (Rel. 36, Last annotation update)  
DE CELLULAR TUMOR ANTIGEN P53.  
GN TP53.  
OS Bos taurus (Bovine), and Bos indicus (zebu).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae;  
OC Bovinae; Bos.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC SPECIES=BOVINE; TISSUE=LIVER;

QY 1 GSRHSHL 9

RESULT 6

ID P53 HUMAN STANDARD; PRT; 393 AA.

AC P04637;

DT 13-AUG-1987 (Rel. 05, Created)

DT 01-MAR-1989 (Rel. 10, Last sequence update)

DT 15-JUL-1999 (Rel. 38, Last annotation update)

DE CELLULAR TUMOR ANTIGEN P53 (PHOSPHOPROTEIN P53).

GN TP53.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

OC Eutheria; Primates; Catarrhini; Hominidae; Homo.

RN [1]

RX MEDLINE; 85230577.

RA ZAKUT-HOURI R., BIENZ-TADMOR B., GIVOL D., OREN M.;

RT "Human p53 cellular tumor antigen: cDNA sequence and expression in

RT COS cells.";

RL EMBO J. 4:1251-1255(1985).

RN [2]

RX MEDLINE; 85267676.

RA HARTON E., WILLIAMSON N.M., RALSTON R., HELFMAN D.M., ADAMS T.E.;

RT "Molecular cloning and in vitro expression of a cDNA clone for human

RT cellular tumor antigen p53.";

RL Mol. Cell. Biol. 5:1601-1610(1985).

RN [3]

RX MEDLINE; 85267676.

RA HARRIS N., BRILL E., SHOHAT O., PROKOCIMER M., WOLF D., ARAI N.,

RA ROTTER V.;

RT "Molecular basis for heterogeneity of the human p53 protein.";

RL Mol. Cell. Biol. 6:4650-4656(1986).

RN [5]

RX MEDLINE; 89108008.

RA BUCHMAN V.L., CHUMAKOV P.M., NINKINA N.N., SAMARINA O.P.,

RA GEORGIEV G.P.;

RT "A variation in the structure of the protein-coding region of the

RT human p53 gene.";

RL Gene 70:245-252(1988).

RN [6]

RX MEDLINE; 85126934.

RA MATLASHIEWSKI G., LAMB P., PIM D., PEACOCK J., CRAWFORD L.,

RA BENCHIMOL S.;

RT "Isolation and characterization of a human p53 cDNA clone: expression

RT of the human p53 gene.";

RL EMBO J. 3:3257-3262(1984).

RN [7]

RX MEDLINE; 90191730.

RA ADDISON C., JENKINS J.R., STURZBECHER H.-W.;

RT "The p53 nuclear localisation signal is structurally linked to a

RT p34cdc2 kinase motif.";

RL Oncogene 5:423-426(1990).

RN [8]

RX MEDLINE; 90280456.

RA BISCHOFF J.R., FRIEDMAN P.N., MARSHAK D.R., PRIVES C., BEACH D.;

RT "Human p53 is phosphorylated by p60-cdc2 and cyclin B-cdc2.";

RL Proc. Natl. Acad. Sci. U.S.A. 87:4766-4770(1990).

RN [9]

RX MEDLINE; 91172186.

RA DE VRIES E.M.G., RICKE D.O., DE VRIES T.N., HARTMANN A., BLASZYK H.,

RA LIAO D., SOUSI T., KOVACH J.S., SOMMER S.S.;

RT "Database of mutations in the p53 and APC tumor suppressor genes

RT designed to facilitate molecular epidemiological analyses.";

RL Hum. Mutat. 7:202-213(1996).

RN [19]

RX MEDLINE; 91153807.

RA SCHEIDTMANN K.H., MUMBY M.C., RUNDELL K., WALTER G.;

RT "Dephosphorylation of simian virus 40 large-T antigen and p53 protein

RT by protein phosphatase 2A: inhibition by small-t antigen.";

RL Mol. Cell. Biol. 11:1996-2003(1991).

RN [10]

RX MEDLINE; 94294808.

RA CLORE G.M., OMICHINSKI J.G., SAKAGUCHI K., ZAMBRANO N., SAKAMOTO H.,

RA APPELLA E., GRONENBORN A.M.;

RT "High-resolution structure of the oligomerization domain of p53 by

RT multidimensional NMR.";

RL Science 265:386-391(1994).

RN [11]

RX MEDLINE; 95292092.

RA LEE W., HARVEY T.S., YIN Y., YAU P., LITCHEFIELD D., ARROWSMITH C.H.;

RT "Solution structure of the tetrameric minimum transforming domain of

RT p53.";

RL Nat. Struct. Biol. 1:877-890(1994).

RN [12]

RX MEDLINE; 98026899.

RA MCCOY M., STAVRIDIS E.S., WATERMAN J.L., WIECZOREK A.M., OPELA S.J.,

RA HALAZONETIS T.D.;

RT "Hydrophobic side-chain size is a determinant of the

RT three-dimensional structure of the p53 oligomerization domain.";

RL EMBO J. 16:6230-6236(1997).

RN [13]

RX MEDLINE; 94294806.

RA CHO Y., GORINA S., JEFFREY P.D., PAVLETICH N.P.;

RT "Crystal structure of a p53 tumor suppressor-DNA complex:

RT understanding tumorigenic mutations.";

RL Science 265:346-355(1994).

RN [14]

RX MEDLINE; 97081030.

RA KUSSIE P.H., GORINA S., MARECHAL V., ELENBAAS B., MOREAU J.,

RA LEVINE A.J., PAVLETICH N.P.;

RT "Structure of the MDM2 oncoprotein bound to the p53 tumor suppressor

RT transactivation domain.";

RL Science 274:948-953(1996).

RN [15]

RX MEDLINE; 97035414.

RA GORINA S., PAVLETICH N.P.;

RT "Structure of the p53 tumor suppressor bound to the ankyrin and SH3

RT domains of 53BP2.";

RL Science 274:1001-1005(1996).

RN [16]

RX MEDLINE; 94090335.

RA HARRIS C.C.;

RT "p53: at the crossroads of molecular carcinogenesis and risk

RT assessment.";

RL Science 262:1980-1981(1993).

RN [17]

RX MEDLINE; 91289156.

RA HOOLSTEIN M., SIDRANSKY D., VOGELSTEIN B., HARRIS C.C.;

RT "p53 mutations in human cancers.";

RL Science 253:49-53(1991).

RN [18]

RX MEDLINE; 96271983.

RA DE VRIES E.M.G., RICKE D.O., DE VRIES T.N., HARTMANN A., BLASZYK H.,

RA LIAO D., SOUSI T., KOVACH J.S., SOMMER S.S.;

RT "Database of mutations in the p53 and APC tumor suppressor genes

RT designed to facilitate molecular epidemiological analyses.";

RL Hum. Mutat. 7:202-213(1996).

RN [19]

RX MEDLINE; 91153807.

KW Nuclear protein; Phosphorylation; Apoptosis.  
FT DOMAIN 1 80 ASP/GLU-RICH (ACIDIC).  
FT DOMAIN 81 150 HYDROPHOBIC.  
FT DOMAIN 319 393 HIGHLY BASIC AND MAY BE INVOLVED IN  
FT INTERACTION WITH DNA.  
FT NUCLEAR LOCALIZATION SIGNAL.  
FT MOD\_RES 392 392 PHOSPHORYLATION (BY SIMILARITY).  
SQ SEQUENCE 393 AA; 43655 MW; 11A9B7F8 CRC32;

Query Match 100.0%; Score 58; DB 1; Length 393;  
Best Local Similarity 100.0%; Pred. No. 1.71e-04;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 361 GSRASHSHL 369  
|||||  
QY 1 GSRASHSHL 9

RESULT 4  
ID P53\_CERAE STANDARD; PRT; 393 AA.  
AC P13481;  
DT 01-JAN-1990 (Rel. 13, Created)  
DT 01-JAN-1990 (Rel. 13, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE CELLULAR TUMOR ANTIGEN P53.  
GN TP53.  
OS Cercopithecus aethiops (Green monkey) (Grivet).  
OC Eukaryota; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Primates; Catarrhini; Cercopithecoidea; Cercopithecoinae;  
OC Chlorocebus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 90045967.  
RA RIGAUDY P., ECKHARDT W.;  
RT "Nucleotide sequence of a cDNA encoding the monkey cellular  
phosphoprotein p53.";  
RL Nucleic Acids Res. 17:8375-8375(1989).  
CC -1- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES  
GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL  
CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN  
TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION, IT IS A  
TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION  
BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF  
THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.  
APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF  
BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2  
EXPRESSION.  
CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
CC -1- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY  
OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED  
IN MANY TYPES OF CANCER.  
CC -1- SIMILARITY: BELONGS TO THE P53 FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; X16384; CAA34420.1; -  
CC PIR; S06594;  
CC HSP; P04637; ISAH.  
CC PROSITE; PS00348; P53; 1.  
CC PFAM; PF00870; P53; 1.  
CC Anti-oncogene; DNA-binding; Transcription regulation; Activator;  
KW Nuclear protein; Phosphorylation; Apoptosis.  
FT DOMAIN 1 68 ASP/GLU-RICH (ACIDIC).  
FT DOMAIN 81 150 HYDROPHOBIC.  
FT DOMAIN 319 393 HIGHLY BASIC AND MAY BE INVOLVED IN  
FT INTERACTION WITH DNA.

FT DOMAIN 311 323 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
FT MOD\_RES 392 392 PHOSPHORYLATION (BY SIMILARITY).  
SQ SEQUENCE 393 AA; 43696 MW; BB7DC62 CRC32;

Query Match 100.0%; Score 58; DB 1; Length 393;  
Best Local Similarity 100.0%; Pred. No. 1.71e-04;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 361 GSRASHSHL 369  
|||||  
QY 1 GSRASHSHL 9

RESULT 5  
ID P53\_MACFA STANDARD; PRT; 393 AA.  
AC P56423;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 15-JUL-1998 (Rel. 36, Last annotation update)  
DE CELLULAR TUMOR ANTIGEN P53.  
GN TP53 OR P53.  
OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Primates; Catarrhini; Cercopithecoidea; Cercopithecoinae;  
OC Macaca.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX KHAM M.A., HANSEN C., WELSH J.A., BENNETT W.P.;  
RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES  
GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL  
CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN  
TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION, IT IS A  
TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION  
BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF  
THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.  
APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF  
BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2  
EXPRESSION.  
CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
CC -1- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY  
OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED  
IN MANY TYPES OF CANCER.  
CC -1- SIMILARITY: BELONGS TO THE P53 FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; U48957; AAB91535.1; -  
CC HSP; P04637; ISAH.  
CC PROSITE; PS00348; P53; 1.  
CC PFAM; PF00870; P53; 1.  
CC Anti-oncogene; DNA-binding; Transcription regulation; Activator;  
KW Nuclear protein; Phosphorylation; Apoptosis.  
FT DOMAIN 1 80 ASP/GLU-RICH (ACIDIC).  
FT DOMAIN 81 150 HYDROPHOBIC.  
FT DOMAIN 319 393 HIGHLY BASIC AND MAY BE INVOLVED IN  
FT INTERACTION WITH DNA.  
FT DOMAIN 311 323 NUCLEAR LOCALIZATION SIGNAL.  
FT MOD\_RES 392 392 PHOSPHORYLATION (BY SIMILARITY).  
SQ SEQUENCE 393 AA; 43678 MW; 2499AC47 CRC32;

Query Match 100.0%; Score 58; DB 1; Length 393;  
Best Local Similarity 100.0%; Pred. No. 1.71e-04;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 361 GSRASHSHL 369  
|||||



CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; AF060514; AAC16909.1; -  
DR EMBL; S77819; AAB42022.1; -  
DR HSSP; P04637; 1YCS.  
DR PROSITE; P500348; P53; 1.  
DR PFAM; PF00870; P53; 1.  
KW Anti-oncogene; DNA-binding; Transcription regulation; Activator;  
KW Nuclear protein; Phosphorylation; Apoptosis.  
FT DOMAIN 1 59 ASP/GLU-RICH (ACIDIC).  
FT DOMAIN 68 137 HYDROPHOBIC.  
FT DOMAIN 307 381 HIGHLY BASIC AND MAY BE INVOLVED IN  
FT DOMAIN 299 311 INTERACTION WITH DNA (BY SIMILARITY).  
FT MOD\_RES 380 385 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
FT MOD\_RES 380 385 PHOSPHORYLATION (BY SIMILARITY).  
SQ SEQUENCE 381 AA; 42486 MW; 70210B63 CRC32;

Query Match 100.0%; Score 58; DB 1; Length 381;

Best Local Similarity 100.0%; Pred. No. 1.71e-04;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 349 GSRHSHSL 357

Qy 1 GSRHSHSL 9

RESULT 2  
ID P53\_FELCA STANDARD; PRT; 386 AA.  
AC P41685;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE CELLULAR TUMOR ANTIGEN P53.  
GN TP53.  
OS Felis silvestris catus (Cat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Carnivora; Fissipedia; Felidae; Felis.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-LYMPH NODE;  
RX MEDLINE; 94333960.  
RA OKUDA M., UEDA A., SAKAI T., OHASHI T., MOMOI Y., YOUN H.Y.,  
RA WATARI T., GOITSUKA R., TSUJIMOTO H., HASEGAWA A.;  
RT "Cloning of feline p53 tumor-suppressor gene and its aberration in  
RT hematopoietic tumors."  
RL Int. J. Cancer 58:602-607(1994).  
RN [2]  
RP SEQUENCE OF 34-354 FROM N.A.  
RX MEDLINE; 94114699.  
RA OKUDA M., UEDA A., MATSUMOTO Y., MOMOI Y., WATARI T., GOITSUKA R.,  
RA O'BRIEN S.J., TSUJIMOTO H., HASEGAWA A.;  
RT "Molecular cloning and chromosomal mapping of feline p53 tumor  
RT suppressor gene."  
RL J. Vet. Med. Sci. 55:801-805(1993).  
CC -!- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES  
CC GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL  
CC CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN  
CC TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION, IT IS A  
CC TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION  
CC BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF  
CC THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.  
CC APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF  
CC BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2  
CC EXPRESSION.  
CC -!- SUBCELLULAR LOCATION: NUCLEAR.  
CC -!- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY  
CC OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED  
CC IN MANY TYPES OF CANCER.  
CC -!- SIMILARITY: BELONGS TO THE P53 FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; D26608; BAA05653.1; -  
DR EMBL; D16460; BAA03927.1; -  
DR HSSP; P04637; 1SAH.  
DR PROSITE; P500348; P53; 1.  
DR PFAM; PF00870; P53; 1.  
KW Anti-oncogene; DNA-binding; Transcription regulation; Activator;  
KW Nuclear protein; Phosphorylation; Apoptosis.  
FT DOMAIN 1 59 ASP/GLU-RICH (ACIDIC).  
FT DOMAIN 304 316 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
FT MOD\_RES 385 385 PHOSPHORYLATION (BY SIMILARITY).  
FT CONFLICT 285 285 K -> R (IN REF. 2).  
SQ SEQUENCE 386 AA; 42692 MW; D6C7132A CRC32;

Query Match 100.0%; Score 58; DB 1; Length 386;

Best Local Similarity 100.0%; Pred. No. 1.71e-04;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 354 GSRHSHSL 362

Qy 1 GSRHSHSL 9

RESULT 3  
ID P53\_MACMU STANDARD; PRT; 393 AA.  
AC P56424;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE CELLULAR TUMOR ANTIGEN P53.  
GN TP53 OR P53  
OS Macaca mulatta (Rhesus macaque).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Primates; Catarrhini; Cercopitheidae; Cercopitheinae;  
OC Macaca.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA KHAN M.A., HANSEN C., WELSH J.A., BENNETT W.P.;  
RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES  
CC GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL  
CC CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN  
CC TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION, IT IS A  
CC TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION  
CC BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF  
CC THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.  
CC APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF  
CC BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2  
CC EXPRESSION.  
CC -!- SUBCELLULAR LOCATION: NUCLEAR.  
CC -!- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY  
CC OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED  
CC IN MANY TYPES OF CANCER.  
CC -!- SIMILARITY: BELONGS TO THE P53 FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; U48956; AAB91534.1; -

DR HSSP; P04637; 1SAH.

DR PROSITE; P500348; P53; 1.

DR PFAM; PF00870; P53; 1.

KW Anti-oncogene; DNA-binding; Transcription regulation; Activator;

\*\*\*\*\*  
M P S R E H  
(TM)  
\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Sat Apr 15 01:13:14 2000; MasPar time 3.07 Seconds  
Tabular output not generated. 87.568 Million cell updates/sec

Title: >US-08-452-843-21  
Description: (1-9) from US08452843.pep  
Perfect Score: 58  
Sequence: 1 GSRHSSHL 9

Scoring table: PAM 150  
Gap 15

Searched: 82229 seqs, 29864866 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: swiss-prot38  
1:swissprot

Statistics: Mean 20.606; Variance 19.595; scale 1.052

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES				
Result No.	Score	Query Match	Length DB ID	Description
1	58	100.0	381 1	P53_CANFA CELLULAR TUMOR ANTIGEN 1.71e-04
2	58	100.0	386 1	P53_FELCA CELLULAR TUMOR ANTIGEN 1.71e-04
3	58	100.0	393 1	P53_WACMU CELLULAR TUMOR ANTIGEN 1.71e-04
4	58	100.0	393 1	P53_CERAE CELLULAR TUMOR ANTIGEN 1.71e-04
5	58	100.0	393 1	P53_MACFA CELLULAR TUMOR ANTIGEN 1.71e-04
6	58	100.0	393 1	P53_HUMAN CELLULAR TUMOR ANTIGEN 1.71e-04
7	51	87.9	382 1	P53_SHEEP CELLULAR TUMOR ANTIGEN 2.01e-02
8	51	87.9	386 1	P53_BOVIN CELLULAR TUMOR ANTIGEN 2.01e-02
9	49	84.5	391 1	P53_RABIT CELLULAR TUMOR ANTIGEN 7.39e-02
10	43	74.1	201 1	ALIAH_HUMAN ALPHA-1-ACID GLYCOPROTEIN 2.98e+00
11	42	72.4	97 1	NULM_STRPU NADH-UBIQUINONE OXIDOREDUCTASE 5.34e+00
12	42	72.4	166 1	SR21_YEAST SIGNAL RECOGNITION PARTICLE 5.34e+00
13	42	72.4	259 1	YPBG_BACSU HYPOTHETICAL 28.6 KD P 5.34e+00
14	42	72.4	390 1	P53_MOUSE CELLULAR TUMOR ANTIGEN 5.34e+00
15	42	72.4	998 1	EPB3_HUMAN EPHRIN TYPE-B RECEPTOR 5.34e+00
16	41	70.7	371 1	WNT1_XENLA WNT-1 PROTEIN PRECURSOR 9.47e+00
17	41	70.7	444 1	YOR3_GLSU HYPOTHETICAL PROTEIN I 9.47e+00
18	41	70.7	533 1	ARSB_HUMAN ARYL SULFATASE B PRECURSOR 9.47e+00
19	41	70.7	687 1	PPK_ECOLI POLYPHOSPHATE KINASE ( 9.47e+00
20	40	69.0	56 1	ET2_CANFA ENDOTHELIN-2 PRECURSOR 1.66e+01
21	40	69.0	97 1	NULM_PARLI NADH-UBIQUINONE OXIDOREDUCTASE 1.66e+01
22	40	69.0	286 1	AS30_HUMAN ADRENAL SPECIFIC 30 KD 1.66e+01
23	40	69.0	323 1	FASA_BOVIN FASL RECEPTOR PRECURSOR 1.66e+01

RESULT ID	P53_CANFA	STANDARD;	PRT;	381 AA.
AC	Q29537			
DT	01-NOV-1997	(Rel. 35, Created)		
DT	15-DEC-1998	(Rel. 37, Last sequence update)		
DT	15-DEC-1998	(Rel. 37, Last annotation update)		
DE	CELLULAR TUMOR ANTIGEN P53.			
GN	TP53 OR P53.			
OS	Canis familiaris (Dog).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;			
OC	Eutheria; Carnivora; Fissipedia; Canidae; Canis.			
RN	[1]			
RP	SEQUENCE FROM N.Y.A.			
RC	TISSUE=LEUKOCYTE;			
RC	MEDLINE; 98178696			
RA	VELDHOFEN N., MILNER J.;			
RT	"Isolation of canine p53 cDNA and detailed characterization of the full length canine p53 protein."			
RL	Oncogene 16:1077-1084(1998).			
RN	[2]			
RP	SEQUENCE OF 25-300 FROM N.A.			
RC	STRAIN=BEAGLE;			
RC	MEDLINE; 95323915.			
RA	KRAEGL S.A., PAZZI K.A., MADEWELL B.R.;			
RT	"Sequence analysis of canine p53 in the region of exons 3-8."			
RL	Cancer Lett. 92:181-186(1995).			
CC	-1- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION, IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES. APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF BAX AND BCL-2 ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2 EXPRESSION.			
CC	-1- SUBCELLULAR LOCATION: NUCLEAR.			
CC	-1- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED IN MANY TYPES OF CANCER.			
CC	-1- SIMILARITY: BELONGS TO THE P53 FAMILY.			
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial			

**THIS PAGE BLANK (USPTO)**

```
QY      2 SRAHSHL 9

RESULT  13
ENTRY   S37627      #type complete
TITLE   protein-tyrosine kinase (EC 2.7.1.112), receptor-type - human
ORGANISM #formal_name Homo sapiens #common_name man
DATE     19-May-1994 #sequence_revision 03-Aug-1995 #text_change
15-Jan-1999

ACCESSIONS S37627
REFERENCE   S37627      #type complete
#authors   Boehme, B.; Holtrich, U.; Wolf, G.; Luzius, H.; Grzeschik,
            K.H.; Strebhardt, K.; Ruebsamen-Waigmann, H.
#journal   Oncogene (1993) 8:2857-2862
#title     PCR mediated detection of a new human
            receptor-tyrosine-kinase, HEK 2.

#accession S37627      preliminary
#status    ##molecule_type mRNA
#residues  1-998 #label BOE
#cross-references EMBL:X75208; NID:G406867; PID:G406868

CLASSIFICATION #superfamily protein-tyrosine kinase, receptor type eph;
                fibronectin type III repeat homology; protein kinase
                homology; SAM homology

KEYWORDS     ATP; phosphotransferase; transmembrane protein
FEATURE      #domain protein kinase homology #label KIN\
631-899      #region protein kinase ATP-binding motif\
639-647      #domain SAM homology #label SAM
922-988

SUMMARY      #length 998 #molecular-weight 110286 #checksum 4450

Query Match 72.4%; Score 42; DB 2; Length 998;
Best Local Similarity 62.5%; Pred. No. 1.27e+01;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 408 PRVHTSHL 415
:|:|:|
QY      2 SRAHSHL 9

RESULT  14
ENTRY   S69625      #type complete
TITLE   hypothetical protein YDR457w - yeast (Saccharomyces
ORGANISM #formal_name Saccharomyces cerevisiae
DATE     22-Aug-1996 #sequence_revision 06-Sep-1996 #text_change
06-Feb-1998

ACCESSIONS S69625
REFERENCE   S69553
#authors   Dietrich, F.S.
#submission submitted to the EMBL Data Library, August 1995
#description The sequence of S. cerevisiae cosmids 9410, 8035, 8166, and
            9787.

#accession S69625
#molecule_type DNA
#residues  1-3268 #label DIE
#cross-references EMBL:U33050; NID:G927726; PID:G927738; MIPS:YDR457w

GENETICS    SGD:TOM1
#gene
#map_position 4R
#cross-references SGD:S0002865; MIPS:YDR457w
SUMMARY      #length 3268 #molecular-weight 374181 #checksum 6577

Query Match 72.4%; Score 42; DB 2; Length 3268;
Best Local Similarity 66.7%; Pred. No. 1.27e+01;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 2194 GSRPRSHL 2202
|||:|:|
QY      1 GSRAHSHL 9

RESULT  15
ENTRY   TVXLTL1      #type complete
TITLE   transforming protein int-1 precursor - African clawed frog
ORGANISM #formal_name Xenopus laevis #common_name African clawed frog
DATE     30-Jun-1991 #sequence_revision 30-Jun-1991 #text_change
26-Feb-1999

ACCESSIONS S02113; S41630
REFERENCE   S02113
#authors   Noordermeer, J.; Meijlink, F.; Verrilizer, P.; Rijsewijk, F.;
            Destree, O.
#journal   Nucleic Acids Res. (1989) 17:11-18
#title     Isolation of the Xenopus homolog of int-1/wingless and
            expression during neurula stages of early development.

#cross-references MUID:89098373
#accession S02113
#molecule_type mRNA
#residues  1-371 #label NOO
#cross-references EMBL:X13138; NID:G65235; PID:G65236
REFERENCE   S41630
#authors   Gao, X.; Kuiken, G.A.; Baarends, W.M.; Koster, J.G.; Destree,
            O.H.J.
#journal   Oncogene (1994) 9:573-581
#title     Characterization of a functional promoter for the Xenopus
            wnt-1 gene in vivo.

#cross-references MUID:94119599
#accession S41630
#molecule_type DNA
#residues  1-37 #label GAO
#cross-references EMBL:X58845

GENETICS     int-1
#gene
CLASSIFICATION #superfamily int-1 transforming protein
KEYWORDS      glycoprotein; oncogene; transforming protein
FEATURE       1-19
20-371
28,261,279,306,317,
360
#binding_site carbohydrate (Asn) (covalent) #status
predicted
SUMMARY      #length 371 #molecular-weight 41125 #checksum 3277

Query Match 70.7%; Score 41; DB 1; Length 371;
Best Local Similarity 55.6%; Pred. No. 2.13e+01;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 267 GSRSDPHL 275
|||:|:|
QY      1 GSRAHSHL 9

Search completed: Sat Apr 15 01:12:57 2000
Job time : 16 secs.
```

```

ORGANISM #formal_name turkey herpesvirus
DATE 30-Sep-1993 #sequence_revision 20-Aug-1994 #text_change
ACCESSIONS JQ2351
REFERENCE JQ2346
#authors Zelnik, V.; Dartell, R.; Audonnet, J.C.; Smith, G.D.;
#journal Riviere, M.; Pastorek, J.; Ross, L.J.N.
#title J. Gen. Virol. (1993) 74:2151-2162
#molecule_type DNA
#accession JQ2351
#residues 1-384 #label ZEL
CLASSIFICATION #superfamily Marek's disease virus glycoprotein D
KEYWORDS glycoprotein; transmembrane protein
FEATURE
1-25 #domain signal sequence #status predicted #label SIG\
26-384 #product glycoprotein D #status predicted #label MAT\
345-360 #domain transmembrane #status predicted #label TM\
123,215,220 #binding_site carbohydrate (Asn) (covalent) #status
predicted
SUMMARY #length 384 #molecular_weight 43857 #checksum 4684
Query Match 72.4%; Score 42; DB 2; Length 384;
Best Local Similarity 85.7%; Pred. No. 1.27e+01;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Db 230 RAHESHL 236
||| |||
QY 3 RAHSSHL 9

RESULT 12
ENTRY DNMS53 #type complete
TITLE cellular tumor antigen p53 - mouse
ALTERNATE_NAMES oncoprotein p53
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE 28-Aug-1985 #sequence_revision 04-Oct-1996 #text_change
12-Feb-1999
ACCESSIONS A22739; S06336; A02684; S38822; S40014; I48703
REFERENCE A22739
#authors Bienz, B.; Zakut-Houri, R.; Givol, D.; Oren, M.
#journal EMBO J. (1984) 3:2179-2183
#cross-references MUID:85027173
#accession A22739
#molecule_type DNA
#residues 1-134, 'V', 136-390 #label BIE
#cross-references GB:X00876; NID:9871420; PID:9871421; GB:X01237;
GB:K01700; NID:953575; PID:953576
REFERENCE S06336
#authors Chumakov, P.M.
#journal Bioorg. Khim. (1987) 13:1691-1694
#title Primary structure of DNA complementary to murine oncoprotein
p53 mRNA.
#cross-references MUID:88221682
#accession S06336
#status not compared with conceptual translation
#molecule_type mRNA
#residues 1-134, 'V', 136-390 #label CHU
REFERENCE A02684
#authors Zakut-Houri, R.; Oren, M.; Bienz, B.; Lavie, V.; Hazum, S.;
Givol, D.
#journal Nature (1983) 306:594-597
#title A single gene and a pseudogene for the cellular tumour
antigen p53
#cross-references MUID:84068204
#accession A02684
#molecule_type mRNA
#residues 1-159, 'H', 161-167, 'G', 169-233, 'I', 235-390 #label ZAK
#cross-references GB:X01237; GB:K01700; NID:953575
REFERENCE S38822
#authors Arai, N.; Nomura, D.; Yokota, K.; Wolf, D.; Brill, E.;
Shohat, O.; Rotter, V.

```

```

#journal Mol. Cell. Biol. (1986) 6:3232-3239
#title Immunologically distinct p53 molecules generated by
alternative splicing.
#cross-references MUID:87064640
#accession S38822
#status preliminary
#molecule_type mRNA
#residues 1-390 #label ARA1
#cross-references EMBL:M13872; NID:g200198; PID:g200199
#accession S38823
#status preliminary
#molecule_type mRNA
#residues 1-167, 'G', 169-233, 'I', 235-390 #label ARA2
#cross-references EMBL:M13873
REFERENCE S40014
#authors Arai, N.; Nomura, D.; Yokota, K.; Wolf, D.; Brill, E.;
Shohat, O.; Rotter, V.
#submission submitted to the EMBL Data Library, July 1988
#accession S40014
#molecule_type mRNA
#residues 1-167, 'G', 169-390 #label ARA3
#cross-references EMBL:M13873; NID:g200200; PID:g200201
REFERENCE I48703
#authors Jenkins, J.R.; Rudge, K.; Redmond, S.; Wade-Evans, A.
#journal Nucleic Acids Res. (1984) 12:5609-5626
#title Cloning and expression analysis of full length mouse cDNA
sequences encoding the transformation associated protein
p53.
#cross-references MUID:84272240
#accession I48703
#status preliminary; translated from GB/EMBL/DBD
#molecule_type mRNA
#residues 1-47, 'R', 49-78, 'QW', 82-390 #label RES
#cross-references EMBL:X00741; NID:g53570; PID:g53571
COMMENT This DNA-binding protein plays an essential role in the regulation
of cell division, as it is required for the transition from phase
G0 to G1 of the cell cycle.
COMMENT The tetramer association region may exhibit a beta-turn,
beta-sheet, beta-turn, alpha-helix motif.
CLASSIFICATION #superfamily cellular tumor antigen p53
KEYWORDS apoptosis; cell division control; DNA binding; homotetramer;
phosphoprotein; transcription regulation; tumor suppressor;
zinc
FEATURE
1-44 #domain transcription activation #status predicted
#label TRA\
16-26 #region conserved region I\
99-289 #domain DNA-binding core #status predicted #label DBC\
108-121 #region L1 loop\
114-139 #region conserved region II\
160-192 #region L2 loop\
168-178 #region conserved region III\
231-252 #region conserved region IV\
233-248 #region L3 loop\
267-283 #region conserved region V\
313-319 #region nuclear location signal\
319-357 #region tetramer association\
7,9,12,18,23,37 #binding_site phosphate (Ser) (covalent) #status
predicted\
173,176,235,239 #binding_site zinc (Cys, His, Cys, Cys) #status
predicted\
312 #binding_site phosphate (Ser) (covalent) (by cdc2
kinase) #status predicted\
389 #binding_site phosphoryl-RNA (Ser) (covalent) #status
predicted
SUMMARY #length 390 #molecular_weight 43458 #checksum 1260
Query Match 72.4%; Score 42; DB 1; Length 390;
Best Local Similarity 87.5%; Pred. No. 1.27e+01;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Db 359 SRAHSSYL 366
|||||

```

```

SUMMARY      phosphorylation; oxidoreductase; respiratory chain
              #length 97 #molecular-weight 10610 #checksum 1423
Query Match   72.4%; Score 42; DB 2; Length 97;
Best Local Similarity 75.0%; Pred. No. 1.27e+01;
Matches      6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 81 SRTHSSNL 88
   11:11111
Qy 2 SRAHSSHL 9

RESULT 9
ENTRY  signal recognition particle protein SRP21 - yeast
TITLE  (Saccharomyces cerevisiae)
ALTERNATE_NAMES
ORGANISM  protein YKL122c; protein YKL527
#Variety  #formal_name Saccharomyces cerevisiae
DATE      12-Mar-1993 #sequence_revision 12-Mar-1993 #text_change
06-Feb-1998
ACCESSIONS S25360; S37950; S51951
REFERENCE  Colleaue, L.; Richard, G.F.; Thierry, A.; Dujon, B.
            Yeast (1992) 8:325-336
            Sequence of a segment of yeast chromosome XI identifies a new
            mitochondrial carrier, a new member of the G protein
            family, and a protein with the PAKK motif of the H1
            histones.
            #cross-references MUID:92383952
            #accession S25360
            #molecule_type DNA
            #residues 1-167 #label COL
            #cross-references EMBL:S44213; NID:g254447; PID:g254451
REFERENCE  J37938
            #authors Jacquier, A.; Legrain, P.; Colleaue, L.; Richard, G.F.;
            Thierry, A.; Dujon, B.
            #submission submitted to the Protein Sequence Database, March 1994
            #accession S37950
            #molecule_type DNA
            #residues 1-167 #label JAC
            #cross-references EMBL:S28122; NID:g486205; PID:g486206; MIPS:YKL122c
            #accession S51616
            #authors Brown, J.D.; Hann, B.C.; Medzihradsky, K.F.; Niwa, M.;
            Burlingame, A.L.; Walter, P.
            #journal EMBO J. (1994) 13:4390-4400
            #title Subunits of the Saccharomyces cerevisiae signal recognition
            particle required for its functional expression.
            #accession S51951
            #molecule_type protein
            #residues 2-15;45-55;75-85;107-120 #label BRO
GENETICS      SGD:SRP21
#gene
#map_position 11L
#cross-references SGD:S0001605; MIPS:YKL122c
SUMMARY      #length 167 #molecular-weight 18425 #checksum 4127
Query Match   72.4%; Score 42; DB 2; Length 167;
Best Local Similarity 71.4%; Pred. No. 1.27e+01;
Matches      5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 44 RTHNSHL 50
   11:11111
Qy 3 RAHSSHL 9

RESULT 10
ENTRY  signal recognition particle protein ypbG - Bacillus subtilis
TITLE  conserved hypothetical protein ypbG - Bacillus subtilis
ORGANISM  #formal_name Bacillus subtilis
DATE      05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change
17-Jul-1998
ACCESSIONS B69933

```

```

REFERENCE
#authors
A69580
Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.;
Alloni, G.; Azevedo, V.; Bertero, M.G.; Bessieres, P.;
Borolin, A.; Borchert, S.; Boriss, R.; Boursier, L.; Brans,
A.; Braun, M.; Brignell, S.C.; Bron, S.; Brouillet, S.;
Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
Choi, S.K.; Codani, J.J.; Connerton, I.F.; Cummings, N.J.;
Daniel, R.A.; Denizot, F.; Devine, K.M.; Duesterhoef, A.;
Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Erington, J.;
Fabret, C.; Ferrari, E.; Foulger, D.; Fritz, C.; Fujita,
M.; Fujita, Y.; Funa, S.; Galizzi, A.; Galleron, N.; Ghim,
S.Y.; Glaser, P.; Goffeau, A.; Golligly, E.J.; Grandi, G.;
Guiseppi, G.; Guy, B.J.; Haga, K.; Halech, J.; Harwood,
C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.;
Hullo, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karamata, D.;
Kasahara, Y.; Klaerr-Blanchard, M.; Klein, C.; Kobayashi,
Y.; Koetter, P.; Konigstein, G.; Krogh, S.; Kumano, M.;
Kurita, K.; Lapidus, A.; Lardinois, S.; Lauber, J.;
Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.;
Maeel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno,
M.; Moestl, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly,
V.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro,
V.; Pohl, T.M.; Portetelle, D.; Porwolik, S.; Prescott,
A.M.; Prescan, E.; Pujic, P.; Purnelle, B.; Rapoport, G.;
Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.;
Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon, E.;
Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.;
Sekowska, A.; Seror, S.J.; Serror, P.; Shin, B.S.; Soldo,
B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Takahashi, H.;
Takemaru, K.; Takeuchi, M.; Tamakoshi, A.; Tanaka, T.;
Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.;
Vandenbol, M.; Vannier, F.; Vassarotti, A.; Viari, T.;
Wambutt, R.; Wedler, E.; Wedler, H.; Weitzenecker, T.;
Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto,
K.; Yata, K.; Yoshida, K.; Yoshikawa, H.F.; Zumstein, E.;
Yoshikawa, H.; Danchin, A.
Nature (1997) 390:249-256
The complete genome sequence of the Gram-positive bacterium
Bacillus subtilis.
#cross-references MUID:98044033
#accession B69933
#status preliminary; nucleic acid sequence not shown;
translation not shown
#molecule_type DNA
#residues 1-259 #label KUN
#cross-references GB:Z99115; GB:Z99116; GB:AL009126; NID:g2634723;
PID:el185567; PID:g2634733; NID:g2634478;
PID:el183743; PID:g2634716
#experimental_source strain 168
COMMENT Although this sequence has motifs characteristic of a variety of
phosphoesterases, a critical active site residue is not
conserved.
GENETICS      ypbG
#gene
CLASSIFICATION #superfamily probable phosphoesterase yaeI; phosphoesterase
core homology
FEATURE
48-114 #domain phosphoesterase core homology #label PEC
SUMMARY      #length 259 #molecular-weight 28560 #checksum 3614
Query Match   72.4%; Score 42; DB 2; Length 259;
Best Local Similarity 50.0%; Pred. No. 1.27e+01;
Matches      4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 68 ARSHAPHL 75
   11:11111
Qy 2 SRAHSSHL 9

RESULT 11
ENTRY  glycoprotein D precursor - turkey herpesvirus
TITLE  glycoprotein D precursor - turkey herpesvirus
ALTERNATE_NAMES ORF 6 protein

```

```

#map_position 8R
KEYWORDS      transmembrane protein
FEATURE
15-31
SUMMARY      #domain transmembrane #status predicted #label TMM
              #length 143 #molecular-weight 16054 #checksum 8917
              79.38; Score 46; DB 2; Length 143;
Query Match  Best Local Similarity 66.7%; Pred. No. 1.45e+00;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 85 GRGHSDHL 93
| | | | |
QY 1 GSRAHSHL 9

RESULT 6
ENTRY  C70892
TITLE  #type complete
        hypothetical protein Rv1061 - Mycobacterium tuberculosis
        (strain H37Rv)
ORGANISM #formal_name Mycobacterium tuberculosis
DATE 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change
C70892
REFERENCE A70500
#authors Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher,
        C.; Harris, D.; Gordon, S.V.; Eigmeier, K.; Gas, S.; Barry
        III, C.E.; Tekala, F.; Badcock, K.; Basham, D.; Brown, D.;
        Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.;
        Felkwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;
        Hornsby, T.; Jagels, K.; Krogh, A.; McLean, J.; Moule, S.;
        Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.;
        Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.;
        Skelton, S.; Squares, S.; Squires, R.; Sulston, J.E.;
        Taylor, K.; Whitehead, S.; Barrell, B.G.
#journal Nature (1998) 393:537-544
#title Deciphering the biology of Mycobacterium tuberculosis from
        the complete genome sequence.
#cross-references MUID:98295987
#accession C70892
#status preliminary; nucleic acid sequence not shown;
        translation not shown
#molecule_type DNA
#residues 1-287 #label COL
#cross-references GB:AL021897; GB:AL123456; NID:g3256022; PID:cl251940;
        PID:g2896698
#experimental_source strain H37Rv
GENETICS
#gene Rv1061
SUMMARY #length 287 #molecular-weight 32037 #checksum 8240
Query Match 77.6%; Score 45; DB 2; Length 287;
Best Local Similarity 85.7%; Pred. No. 2.52e+00;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 214 RAHSTHL 220
| | | | |
QY 3 RAHSHL 9

RESULT 7
ENTRY  OMHU2
TITLE  #type complete
        alpha-1-acid glycoprotein 2 precursor - human
        alpha-1-acid glycoprotein B; orosomucoid 2
        #formal_name Homo sapiens #common_name man
        31-Mar-1992 #sequence_revision 07-Jun-1996 #text_change
        08-May-1998
ACCESSIONS JT0326; B28346
REFERENCE JT0326
#authors Merritt, C.M.; Board, P.G.
#journal Gene (1988) 66:97-106
#title Structure and characterisation of a duplicated human alpha 1
        acid-glycoprotein gene.
#cross-references MUID:88329732

```

```

#accession JT0326
#molecule_type DNA
#residues 1-201 #label MER
#cross-references GB:M21540; NID:g177839; PID:g177840
REFERENCE A28346
#authors Dente, L.; Pizza, M.G.; Metapalu, A.; Cortese, R.
#journal EMBO J. (1987) 6:2289-2296
#title Structure and expression of the genes coding for human
        alpha-1-acid glycoprotein.
#cross-references MUID:88029318
#accession B28346
#molecule_type DNA
#residues 1-118, 'N', 120-201 #label DEN
#cross-references GB:X06674
COMMENT Alpha-1-AGP, synthesized in the liver and leucocytes, appears to
        function in modulating the activity of the immune system during
        the acute-phase reaction.
COMMENT See also PIR:OMHU1.
GENETICS
#gene GDB:ORM2
#cross-references GDB:120251; OMIM:138610
#map_position 9q32-9q32
#introns 38/3; 86/2; 110/1; 146/1; 180/3
CLASSIFICATION #superfamily lipocalin; lipocalin homology
KEYWORDS acute phase; glycoprotein; leukocyte; liver; plasma;
        pyroglyutamic acid
FEATURE
1-18 #domain signal sequence #status predicted #label SIG\
19-201 #product alpha-1-acid glycoprotein 2 #status predicted
        #label MAT\
34-183 #domain lipocalin homology #label LIP\
19 #modified_site pyrrolidone carboxylic acid (Gln) (in
        mature form) #status predicted\
23-165,90-183 #disulfide_bonds #status predicted\
33,56,72,93,103 #binding_site carbohydrate (Asn) (covalent) #status
        predicted
SUMMARY #length 201 #molecular-weight 23602 #checksum 4589
Query Match 74.1%; Score 43; DB 1; Length 201;
Best Local Similarity 55.6%; Pred. No. 7.45e+00;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 111 GREHVAHL 119
| | | | |
QY 1 GSRAHSHL 9

RESULT 8
ENTRY  S01502
TITLE  #type complete
        NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 4L - sea
        urchin (Strongylocentrotus purpuratus) mitochondrion (SGC8)
        #formal_name Strongylocentrotus purpuratus purpuratus
        #common_name purple urchin
        01-Dec-1989 #sequence_revision 01-Dec-1989 #text_change
        09-Sep-1994
ACCESSIONS S01502
REFERENCE S01499
#authors Jacobs, H.T.; Elliott, D.J.; Math, V.B.; Farquharson, A.
        J. Mol. Biol. (1988) 202:185-217
#journal Nucleotide sequence and gene organization of sea urchin
        mitochondrial DNA.
#title mitochondrion DNA.
#cross-references MUID:89011951
#accession S01502
#molecule_type DNA
#residues 1-97 #label JAC
#cross-references EMBL:X12631
GENETICS
#gene nd4L
#genome mitochondrion
#genetic_code SGC8
#start_codon ATC
#classification #superfamily NADH dehydrogenase (ubiquinone) chain 4L
        membrane-associated complex; mitochondrion; NAD; oxidative
        phosphorylation
KEYWORDS

```

#authors Yamada, Y.; Yoshida, T.; Hayashi, K.; Sekiya, T.; Yokota, J.; Hirohashi, S.; Nakatani, K.; Nakano, H.; Sugimura, T.; Terada, M.  
#journal Cancer Res. (1991) 51:5800-5805  
#title p53 gene mutations in gastric cancer metastases and in gastric cancer cell lines derived from metastases.  
#cross-references MUID:92034678  
#accession R44905

...  
Note: remainder of annotations omitted.

Query Match 100.0%; Score 58; DB 1; Length 393;  
Best Local Similarity 100.0%; Pred. No. 1.12e-03;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 361 GSAHSHSL 369  
|||||  
Qy 1 GSAHSHSL 9

RESULT 2  
ENTRY S06594 #type complete  
TITLE cellular tumor antigen p53 - green monkey  
ORGANISM #formal\_name Cercopithecus aethiops #common\_name green monkey, grivet  
DATE 28-Feb-1990 #sequence\_revision 28-Feb-1990 #text\_change 08-Sep-1997  
ACCESSIONS S06594  
REFERENCE S06594  
#authors Rigaudy, P.; Eckhart, W.  
#journal Nucleic Acids Res. (1989) 17:8375  
#title Nucleotide sequence of a cDNA encoding the monkey cellular phosphoprotein p53.  
#cross-references MUID:90045967  
#accession S06594  
#molecule\_type mRNA  
#residues 1-393 #label RIG  
#cross-references EMBL:X16384; NID:g22795; PID:g22796  
CLASSIFICATION #superfamily cellular tumor antigen p53  
KEYWORDS apoptosis; cell division control; DNA binding; homotetramer; nucleus; phosphoprotein; transcription regulation; tumor suppressor; zinc

FEATURE 176,179,238,242 #binding\_site zinc (Cys, His, Cys) #status predicted  
392 #binding\_site phosphoryl-RNA (Ser) (covalent) #status predicted

SUMMARY #length 393 #molecular-weight 43696 #checksum 4263  
Query Match 100.0%; Score 58; DB 2; Length 393;  
Best Local Similarity 100.0%; Pred. No. 1.12e-03;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 361 GSAHSHSL 369  
|||||  
Qy 1 GSAHSHSL 9

RESULT 3  
ENTRY S51648 #type complete  
TITLE cellular tumor antigen p53 - bovine  
ALTERNATE\_NAMES tumor-suppressor protein p53  
ORGANISM #formal\_name Bos primigenius taurus #common\_name cattle  
DATE 07-May-1995 #sequence\_revision 01-Sep-1995 #text\_change 08-Sep-1997  
ACCESSIONS S51648  
REFERENCE S51648  
#authors Dequiedt, F.; Willems, L.; Burny, A.; Kettmann, R.  
#submission submitted to the EMBL Data Library, September 1994  
#description Nucleotide sequence of the ovine p53 tumor-suppressor gene cDNA and its genomic organisation.  
#accession S51648  
... #status preliminary

##molecule\_type mRNA  
##residues 1-386 #label DEQ  
##cross-references EMBL:X81704; NID:g602332; PID:g602333  
CLASSIFICATION #superfamily cellular tumor antigen p53  
KEYWORDS apoptosis; cell division control; DNA binding; homotetramer; phosphoprotein; transcription regulation; tumor suppressor; zinc  
FEATURE 168,171,231,235 #binding\_site zinc (Cys, His, Cys) #status predicted  
385 #binding\_site phosphoryl-RNA (Ser) (covalent) #status predicted  
SUMMARY #length 386 #molecular-weight 43255 #checksum 7025  
Query Match 87.9%; Score 51; DB 2; Length 386;  
Best Local Similarity 100.0%; Pred. No. 8.13e-02;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 355 SRAHSHSL 362  
|||||  
Qy 2 SRAHSHSL 9

RESULT 4  
ENTRY JC6193 #type complete  
TITLE tumor suppressor p53 - rabbit  
ORGANISM #formal\_name Oryctolagus cuniculus #common\_name domestic rabbit  
DATE 11-Apr-1997 #sequence\_revision 09-May-1997 #text\_change 17-Mar-1999  
ACCESSIONS JC6193  
REFERENCE JC6193  
#authors Le Goas, F.; May, P.; Ronco, P.; de Fromental, C.C.  
#journal Gene (1997) 185:169-173  
#title cDNA cloning and immunological characterization of rabbit p53.

##cross-references MUID:97208869  
#accession JC6193  
#molecule\_type mRNA  
#residues 1-391 #label LEA  
#cross-references EMBL:X90592; NID:g1532043; PID:g194562; PID:g1532044  
GENETICS

#gene p53  
CLASSIFICATION #superfamily cellular tumor antigen p53  
KEYWORDS tumor  
SUMMARY #length 391 #molecular-weight 43435 #checksum 4367

Query Match 84.5%; Score 49; DB 2; Length 391;  
Best Local Similarity 88.9%; Pred. No. 2.62e-01;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 359 GSAHSHSL 367  
|||||  
Qy 1 GSAHSHSL 9

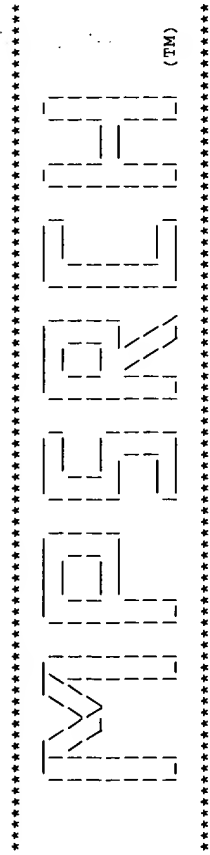
RESULT 5  
ENTRY S52595 #type complete  
TITLE probable membrane protein YHR056w-a - yeast (Saccharomyces cerevisiae)  
ORGANISM #formal\_name Saccharomyces cerevisiae  
DATE 05-May-1995 #sequence\_revision 19-Oct-1995 #text\_change 21-Nov-1997  
ACCESSIONS S52595  
REFERENCE S46729  
#authors Du, Z.  
#submission submitted to the EMBL Data Library, May 1994  
#description The sequence of S. cerevisiae cosmid 8025.  
#accession S52595  
#molecule\_type DNA  
#residues 1-143 #label DUZ  
#cross-references EMBL:U00061; MIPS:YHR056w-a

GENETICS



##cross-references EMBL:X54156; NID:g35213; PID:g35214  
REFERENCE S42669  
#authors Matlashewski, G.; Lamb, P.; Pim, D.; Peacock, J.; Crawford, L.; Benchimol, S.  
#journal EMBO J. (1994) 3:3257-3262  
#title Isolation and characterization of a human p53 cDNA clone: expression of the human p53 gene.  
##cross-references MUID:85126934  
#accession S42669  
##molecule\_type mRNA  
##residues 101-393 ##label MK11  
##cross-references EMBL:X01405; NID:g35215; PID:g642241  
REFERENCE A22837  
#authors Zakut-Houri, R.; Bienz-Tadmor, B.; Givol, D.; Oren, M.  
#journal EMBO J. (1995) 4:1251-1255  
#title Human p53 cellular tumor antigen: cDNA sequence and expression in COS cells.  
##cross-references MUID:85230577  
#accession A22837  
##molecule\_type mRNA  
##residues 1-71, 'P', 73-393 ##label ZAK  
##cross-references EMBL:X02469; EMBL:M60950; NID:g35209; PID:g35210  
REFERENCE A55060  
#authors Harlow, E.; Williamson, N.M.; Ralston, R.; Helfman, D.M.; Adams, T.E.  
#journal Mol. Cell. Biol. (1985) 5:1601-1610  
#title Molecular cloning and in vitro expression of a cDNA clone for human cellular tumor antigen p53.  
##cross-references MUID:85267676  
#accession A55060  
##molecule\_type mRNA  
##residues 1-71, 'P', 73-272, 'H', 274-393 ##label HAR  
##cross-references GB:X03199; NID:g189478; PID:g189479  
##experimental\_source clone PR4-2, cell line A431  
REFERENCE A93086  
#authors Harris, M.; Brill, E.; Shohat, O.; Prokocimer, M.; Wolf, D.; Arai, N.; Rotter, V.  
#journal Mol. Cell. Biol. (1986) 6:4650-4656  
#title Molecular basis for heterogeneity of the human p53 protein.  
##cross-references MUID:87089826  
#accession A25397  
##molecule\_type mRNA  
##residues 1-78, 'T', 80-393 ##label HAR1  
##cross-references EMBL:M14694; NID:g339813; PID:g339814  
##experimental\_source clone p53-H-1, transformed hybridoma SV-80 cell line  
#accession B25397  
##molecule\_type mRNA  
##residues 1-71, 'P', 73-78, 'T', 80-393 ##label HAR2  
##cross-references EMBL:M14695; NID:g339815; PID:g339816  
##experimental\_source clone p53-H-19, transformed hybridoma SV-80 cell line  
REFERENCE S42452  
#authors Matlashewski, G.J.; Tuck, S.; Pim, D.; Lamb, P.; Schneider, J.; Crawford, L.V.  
#journal Mol. Cell. Biol. (1987) 7:961-963  
#title Primary structure polymorphism at amino acid residue 72 of human p53.  
##cross-references MUID:87144273  
#accession S42452  
##molecule\_type mRNA; DNA  
##residues 66-71, 'P', 73-79 ##label MK12  
##experimental\_source clone lambda C113  
##note 72-Cys was also found, and appears to represent a polymorphism  
#accession S42453  
##molecule\_type mRNA; DNA  
##residues 66-79 ##label MK13  
##experimental\_source clone J6K  
REFERENCE I38082  
#authors Farrell, P.J.; Allan, G.J.; Shanahan, F.; Vousden, K.H.; Crook, T.  
#journal EMBO J. (1991) 10:2879-2887

#title p53 is frequently mutated in Burkitt's lymphoma cell lines.  
##cross-references MUID:92007731  
#accession I38082  
##status translated from GB/EMBL/DBJ  
##molecule\_type mRNA  
##residues 1-189, 'LLSILSEWKEICVSIWMTETLFDIVWCPMSRLRLALT', 'VPSPTTTCVTVPANAA' ##label F01  
##cross-references EMBL:X60010; NID:g506432; PID:g506433  
##note Deletion of a C nucleotide causes a frameshift at position 566  
#accession I38083  
##status translated from GB/EMBL/DBJ  
##molecule\_type mRNA  
##residues 1-192, 'R', 194-393 ##label F02  
##cross-references EMBL:X60011; NID:g506434; PID:g506435  
#accession I38084  
##status translated from GB/EMBL/DBJ  
##molecule\_type mRNA  
##residues 1-393 ##label F03  
##cross-references EMBL:X60012; NID:g506436; PID:g506437  
#accession I38085  
##status translated from GB/EMBL/DBJ  
##molecule\_type mRNA  
##residues 1-245, 'T', 247-393 ##label F04  
##cross-references EMBL:X60013; NID:g506438; PID:g506439  
#accession I38086  
##status translated from GB/EMBL/DBJ  
##molecule\_type mRNA  
##residues 1-236, 'I', 238-393 ##label F05  
##cross-references EMBL:X60014; NID:g506440; PID:g506441  
#accession I38087  
##status translated from GB/EMBL/DBJ  
##molecule\_type mRNA  
##residues 1-247, 'Q', 249-393 ##label F06  
##cross-references EMBL:X60015; NID:g506442; PID:g506443  
#accession I38088  
##status translated from GB/EMBL/DBJ  
##molecule\_type mRNA  
##residues 1-71, 'P', 73-237, 'Y', 239-393 ##label F07  
##cross-references EMBL:X60016; NID:g506444; PID:g506445  
#accession I38089  
##status translated from GB/EMBL/DBJ  
##molecule\_type mRNA  
##residues 1-247, 'Q', 249-393 ##label F08  
##cross-references EMBL:X60017; NID:g506446; PID:g506447  
#accession I38090  
##status translated from GB/EMBL/DBJ  
##molecule\_type mRNA  
##residues 1-71, 'P', 73-162, 'H', 164-393 ##label F09  
##cross-references EMBL:X60018; NID:g506448; PID:g506449  
#accession I38091  
##status translated from GB/EMBL/DBJ  
##molecule\_type mRNA  
##residues 1-212, 'Q', 214-393 ##label F10  
##cross-references EMBL:X60019; NID:g506450; PID:g506451  
#accession I38092  
##status translated from GB/EMBL/DBJ  
##molecule\_type mRNA  
##residues 1-253, 'D', 255-393 ##label F11  
##cross-references EMBL:X60020; NID:g506452; PID:g506453  
##note all sequences submitted to the EMBL/GenBank/DBJ databases June 1991  
REFERENCE I38093  
#authors Futreal, P.A.; Barrett, J.C.; Wiseman, R.W.  
#journal Nucleic Acids Res. (1991) 19:6977  
#title An Alu polymorphism intragenic to the TP53 gene.  
##cross-references MUID:92107726  
#accession I38093  
##status translated from GB/EMBL/DBJ  
##molecule\_type DNA  
##residues 1-393 ##label FUT  
##cross-references EMBL:X54156; NID:g35213; PID:g35214  
REFERENCE A44905



Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Sat Apr 15 01:12:41 2000; MasPar time 3.18 Seconds  
Tabular output not generated. 113.579 Million cell updates/sec

Title: >US-08-452-843-21  
Description: (1-9) from US08452843.pep  
Perfect Score: 58  
Sequence: 1 GSAHSSHL 9

Scoring table: PAM 150  
Gap 15

Searched: 122810 seqs, 40068593 residues

Post-processing: Minimum Watch 0%  
Listing first 45 summaries

Database: p1r62  
1:p1r1 2:p1r2 3:p1r3 4:p1r4

Statistics: Mean 20.004; Variance 21.623; scale 0.925

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES							
Result No.	Score	Query Match	Length	ID	Description	Pred. No.	
1	58	100.0	393	1	cellular tumor antigen	1.12e-03	
2	58	100.0	393	2	cellular tumor antigen	1.12e-03	
3	51	87.9	386	2	cellular tumor antigen	8.13e-02	
4	49	84.5	391	2	tumor suppressor p53	2.62e-01	
5	46	79.3	143	2	probable membrane protein	1.45e+00	
6	45	77.6	287	2	hypothetical protein	2.52e+00	
7	43	74.1	201	1	alpha-1-acid glycoprotein	7.45e+00	
8	42	72.4	97	2	NADH dehydrogenase (u	1.27e+01	
9	42	72.4	167	2	signal recognition pa	1.27e+01	
10	42	72.4	259	2	conserved hypothetical	1.27e+01	
11	42	72.4	384	2	glycoprotein D precu	1.27e+01	
12	42	72.4	390	1	cellular tumor antigen	1.27e+01	
13	42	72.4	998	2	protein-tyrosine kina	1.27e+01	
14	42	72.4	3268	2	hypothetical protein	1.27e+01	
15	41	70.7	371	1	transforming protein	2.13e+01	
16	41	70.7	533	1	N-acetylgalactosami	2.13e+01	
17	41	70.7	688	2	polyphosphate kinase	2.13e+01	
18	41	70.7	1018	2	alpha-D-mannosidase (	2.13e+01	
19	41	70.7	1534	2	ribosome receptor, 18	2.13e+01	
20	40	69.0	77	2	endothelin 2 precursor	3.55e+01	
21	40	69.0	97	2	NADH dehydrogenase (u	3.55e+01	
22	40	69.0	151	2	transforming protein	3.55e+01	
23	40	69.0	287	2	homeotic protein p62	3.55e+01	

24	40	69.0	414	2	B64033	hypothetical protein	3.55e+01
25	40	69.0	446	2	S59646	hypothetical protein	3.55e+01
26	40	69.0	456	2	S6080	UDP-N-acetylglucosami	3.55e+01
27	40	69.0	533	2	S18539	actVA-1 protein - Str	3.55e+01
28	40	69.0	558	2	S39621	DNA-directed DNA poly	3.55e+01
29	40	69.0	600	2	B46642	DNA-directed DNA poly	3.55e+01
30	40	69.0	621	2	S73155	hypothetical protein	3.55e+01
31	40	69.0	959	2	S61155	hypothetical protein	3.55e+01
32	40	69.0	1032	2	I38510	neuronal kinesin heav	3.55e+01
33	40	69.0	1379	2	S01254	hepatocyte growth fac	3.55e+01
34	40	69.0	1390	1	TVHUME	hepatocyte growth fac	3.55e+01
35	40	69.0	1748	2	JN0786	integrin beta-4 chain	3.55e+01
36	40	69.0	1807	2	JC6319	integrin beta-4 chain	3.55e+01
37	39	67.2	282	2	H71369	conserved hypothetical	5.87e+01
38	39	67.2	386	2	I84612	sensory epithelia neu	5.87e+01
39	39	67.2	437	2	A70587	hypothetical protein	5.87e+01
40	39	67.2	457	2	JC6026	ADP-ribosyltransferas	5.87e+01
41	39	67.2	1070	2	S19686	alpha-glucosidase (EC	5.87e+01
42	39	67.2	1706	2	I84499	zinc finger protein R	5.87e+01
43	39	67.2	3461	2	S58870	reelin precursor - mo	5.87e+01
44	39	67.2	4957	2	T03455	ALR protein - human	5.87e+01
45	39	67.2	5262	2	T03454	ALR protein - human	5.87e+01

ALIGNMENTS

RESULT 1  
ENTRY DNHU53 #type complete  
TITLE cellular tumor antigen p53 - human  
ALTERNATE\_NAMES cellular phosphoprotein p53; oncoprotein p53; transformation suppressor p53; tumor suppressor p53  
ORGANISM #formal\_name Homo sapiens #common\_name man  
DATE 05-Oct-1988 #sequence\_revision 18-Nov-1994 #text\_change 26-Feb-1999  
ACCESSIONS A25224; A43073; J70436; S42669; A22837; A55060; A25397; B25397; S42452; S42453; I38082; I38083; I38084; I38085; I38086; I38087; I38088; I38089; I38090; I38091; I38092; I38093; A44905; I58354; I78850; I52681; S60153  
REFERENCE A25224  
#authors Lamb, P.; Crawford, L.  
#journal Mol. Cell. Biol. (1986) 6:1379-1385  
#title Characterization of the human p53 gene.  
#cross-references MIM:87064416  
#accession A25224  
#molecule\_type DNA  
#residues 1-393 #label LAM  
#cross-references EMBL:X01405; GB:M13121; GB:N00032; NID:g189460; PID:g386994

REFERENCE J70436  
#authors Buchman, V.L.; Chumakov, P.M.; Ninkina, N.N.; Samarina, O.P.; Georgiev, G.P.  
#journal Gene (1988) 70:245-252  
#title A variation in the structure of the protein-coding region of the human p53 gene.  
#cross-references MIM:89108008  
#accession A43073  
#molecule\_type DNA  
#residues 1-393 #label BUC1  
#cross-references EMBL:M22898; NID:g189474  
#note this 72-Arg allele appears to be about 5 times more frequent than the 72-Pro allele  
#accession J70436  
#molecule\_type DNA  
#residues 1-71, 'P', 73-393 #label BUC2  
#cross-references EMBL:M22898; NID:g189474; PID:g189476  
#note this 72-Pro allele was found in both normal and malignant cell lines  
REFERENCE S40773  
#authors Chumakov, P.M.; Almazov, V.P.; Jenkins, J.R.  
#submission submitted to the EMBL Data Library, August 1990  
#accession S40773  
#molecule\_type DNA  
#residues 1-393 #label CHU

FT	Misc_difference 341	
FT	/label= Ile, Leu, Phe or Val	
FT	/note= "wild-type Phe can be replaced by any of the	
FT	above amino acids"	
FT		
FT	Misc_difference 344	
FT	/label= Ala, Ile, Leu, Phe, Trp or Tyr	
FT	/note= "wild-type Leu can be replaced by any of the	
FT	above amino acids"	
FT		
FT	WO9831703-A1.	
PN	23-JUL-1998.	
PD	15-JAN-1998; U00853.	
PF	17-JAN-1997; US-035458.	
PR	(WIST-) WISTAR INST ANATOMY & BIOLOGY.	
PA	Halazonetis TD;	
PI	WPI; 98-414033/35.	
DR	Altering three-dimensional structure of protein without denaturing	
PT	it - by replacing large hydrophobic amino acids with small ones, or	
PT	vice versa, used to, e.g. produce p53 proteins with altered	
PT	oligomerisation properties	
PT	Example 1; Pages 38-39; 49pp; English.	
PS	This represents a p53 mutant used to demonstrate the method of invention	
CC	of altering the three-dimensional (3D) structure of a protein, without	
CC	denaturing the protein. The method comprises identifying hydrophobic	
CC	residues in the protein and classifying the residues as large or small	
CC	according to size of the side chain. Mutants in which the hydrophobic	
CC	residues have been substituted are produced and they can be analysed for	
CC	a change in the 3D structure. The method is applied to non-linear	
CC	proteins having a hydrophobic core. A p53 oligomerisation domain can be	
CC	altered by this method. Vectors containing nucleic acids encoding a p53	
CC	oligomerisation domain and fusion proteins of the p53 oligomerisation	
CC	domain with a heterologous domain are useful as therapeutic and	
CC	diagnostic agents, in biotechnology and other industrial applications.	
CC	Typically altered p53 is useful as tumour suppressor, to induce apoptosis	
CC	in proliferating lymphocytes, to prevent transplant rejection, to treat	
CC	autoimmune diseases such as systemic lupus erythematosus and rheumatoid	
CC	arthritis, and to suppress proliferation in cases of psoriasis,	
CC	atherosclerosis and arterial restenosis. They are also used to diagnose	
CC	diseases associated with p53 and abnormal cell proliferation. Typical	
CC	heterologous proteins are antibodies against Fos or Jun, soluble	
CC	interleukin-2 receptor complex (for screening drugs that bind the native	
CC	receptor and as therapeutic decoys) and transmembrane receptors. The	
CC	p53 oligomerisation domain may also be used to induce dimerisation of	
CC	DNA-binding proteins, especially c-myc. Altered proteins retain	
CC	biological function but have better stability, binding and lower	
CC	molecular weight (so enter cells more easily). When expressed from a	
CC	gene therapy vector, the altered p53 will not be sequestered into	
CC	inactive heterodimers by mutant p53 present in the cells.	
CC	Sequence 393 AA;	
SC		
Query Match	100.0%; Score 58; DB 1; Length 393;	
Best Local Similarity	100.0%; Pred. No. 7,32e-01;	
Matches	9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
Db	361 GSRASHSHL 369	
QY	1 GSRASHSHL 9	
RESULT	15	
ID	W69718 standard; protein; 393 AA.	
AC	W69718;	
DE	21-OCT-1998 (first entry)	
DT	Human p53 used in coupled proteins and fusion proteins.	
KW	Herpesviral VP22 protein; HSV-1; coupled protein; fusion protein;	
KW	human p53; cell cycle; suicide protein; cytotoxic; antigenic;	
KW	microbial; viral; tumour; antigen; immunomodulating; therapeutic.	
OS	Homo sapiens.	
PN	WO9832866-A1.	
PD	30-JUL-1998.	
PF	23-JAN-1998; G00207.	
PR	01-AUG-1997; GB-016398.	
PR	23-JAN-1997; GB-001363.	
PA	(CUR-) CURIE CANCER CARE MARIE.	

```
DR WPI; 99-120883/10.
DR N-PSDB; X28601.
PT New identified mutations in p53 gene - at nucleotide positions 489,
PT 537 and 1279, used to develop products for the detection of tumours,
PT particularly colorectal cancer
PT Disclosure; Pages 38-39; 44pp; English.
CC This is the amino acid sequence of the p53 protein used in the
CC method of the invention to create mutants. The products and methods
CC can be used for identifying p53 mutations which are indicative of
CC tumours, particularly colorectal cancer. They can also be used for
CC producing transgenic animals which can be used in drug screening
CC assays.
SQ Sequence 393 AA;
  Query Match 100.0%; Score 58; DB 1; Length 393;
  Best Local Similarity 100.0%; Pred. No. 7.32e-01;
  Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 361 GSRASHSHL 369
  Y 1 GSRASHSHL 9
  |||||
  Y 1 GSRASHSHL 9

RESULT 13
ID W84270 standard; Protein; 393 AA.
AC W84270;
DE 13-APR-1999 (first entry)
DE Human p53 protein.
KW Ataxia telangiectasia; ATM protein; assay; interaction; kinase activity;
KW p53; screening; ATR.
OS Homo sapiens.
PN GB2327498-A.
PD 27-JAN-1999.
PF 16-JUL-1998; 015423.
PR 16-JUL-1997; GB-014971.
PA (CANC-) CANCER RES CAMPAIGN TECHNOLOGY.
PI Jackson SP, Lakin ND, Smith GCM;
DR N-PSDB; X04533.
PT Assay method for compounds modulating the interaction of ATM and p53
PT - useful for the treatment of e.g. cancer, immunosuppression and HIV
PT infections and for the purification of the proteins ATM and ATR
PS Disclosure; Fig 7a; 124pp; English.
CC The present sequence represents a human p53 protein. The protein is
CC used in the assay of the invention. The specification describes an
CC assay method for a compound able to modulate the interaction between
CC ATM or a protein having an associated kinase activity and p53 or a
CC protein having homologous phosphorylation sites. The assay comprises
CC contacting a peptide fragment ATM with a relevant fragment of p53
CC and a test compound, and determining the interaction or binding
CC between the substances and the test compound. The assay method is
CC useful for screening for compounds able to modulate the interaction
CC between ATM and p53. The screened agents, peptide fragments and
CC nucleic acids are useful for therapy involving modulating ATM action
CC e.g. in the treatment of cancer, immunosuppression or HIV infections by
CC modulating phosphorylation of p53 by ATM, and for purifying the proteins
CC ATM and ATR.
SQ Sequence 393 AA;
  Query Match 100.0%; Score 58; DB 1; Length 393;
  Best Local Similarity 100.0%; Pred. No. 7.32e-01;
  Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 361 GSRASHSHL 369
  Y 1 GSRASHSHL 9
  |||||
  Y 1 GSRASHSHL 9

RESULT 14
ID W69219 standard; Protein; 393 AA.
AC W69219;
DE 19-OCT-1998 (first entry)
DE Human p53 mutant 2.
KW Hydrophobic; p53 protein; mutant; oligomerisation domain; dimerisation;
KW therapeutic agent; biotechnology; tumour suppressor; apoptosis; human;
KW transplant rejection; autoimmune disease; systemic lupus erythematosus;
KW rheumatoid arthritis; psoriasis; atherosclerosis; arterial restenosis;
KW abnormal cell proliferation; interleukin-2 receptor complex;
KW three-dimensional structure; 3D structure; transmembrane receptor.
OS Homo sapiens.
FT Key Location/Qualifiers
FT Misc_difference 340
FT /label= M340K
FT /note= "wild-type Met is replaced by Lys"
```

PD 06-FEB-1997.  
PF 17-JUL-1996; F01111.  
PR 19-JUL-1995; FR-008729.  
PA (RHON ) RHONE POULENC RORER SA.  
PI Bracco L, Conseiller E;  
DR WPI: 97-132633/12.  
PT New p53 variants e.g. with oligomerisation domain replaced by  
PT leucine zipper - useful for treating hyper-proliferative disorders,  
PT esp. cancer and restenosis  
PS Claim 37; Page -: 133pp; French.  
CC Claimed variants of protein p53 have at least part of the  
CC oligomerisation domain deleted and replaced by a leucine zipper  
CC domain. The mutants preferably also have at least part of the p53  
CC transactivation domain (amino acids 1-74) deleted and replaced by  
CC the domain 325-393 of p53. The present sequence is that of a  
CC specifically claimed p53 variant designated 393-325H and comprising  
CC the 325-393 domain, amino acids 75-325 of human wild-type p53 (but with  
CC Arg182 replaced by His) and a leucine zipper domain at the C-terminal.  
CC The p53 variants are more active and more stable tumour suppressors  
CC and apoptosis-inducing agents than wild-type p53 and are active where  
CC the wild-type protein is not, i.e. they are not inactivated by dominant  
CC negative or oncogenic mutants, nor by other cellular proteins (because  
CC the leucine zipper domain prevents formation of inactive mixed  
CC oligomers).  
CC (Note: this sequence does not appear in the specification and has  
CC been produced by modifying the given sequence of variant 393-325).  
SQ Sequence 353 AA;  
Query Match 100.0%; Score 58; DB 1; Length 353;  
Best Local Similarity 100.0%; Pred. No. 7.32e-01;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Db 38 GSRHSSHL 46  
QY 1 GSRHSSHL 9  
RESULT 9  
ID W28493;  
AC W28493 standard; Protein; 353 AA.  
DE 25-NOV-1997 (first entry)  
DT Human p53 protein variant 393-325 encoded by pEC177.  
KW Leucine zipper domain; LZB; oligomerisation domain; mutant; mutein;  
KW substitution; replacement; transactivation; viral protein VP16; HSV;  
KW anti-oncogene; hyperproliferation; cancer; restenosis;  
KW tumour suppression; apoptosis.  
OS Homo sapiens.  
OS Synthetic.  
PN WO9704092-A1.  
PD 06-FEB-1997.  
PF 17-JUL-1996; F01111.  
PR 19-JUL-1995; FR-008729.  
PA (RHON ) RHONE POULENC RORER SA.  
PI Bracco L, Conseiller E;  
DR WPI: 97-132633/12.  
DR N-PSDB: T86222.  
PT New p53 variants e.g. with oligomerisation domain replaced by  
PT leucine zipper - useful for treating hyper-proliferative disorders,  
PT esp. cancer and restenosis  
PS Claim 37; Pages 90-92; 133pp; French.  
CC Claimed variants of protein p53 have at least part of the  
CC oligomerisation domain deleted and replaced by a leucine zipper  
CC domain. The mutants preferably also have at least part of the p53  
CC transactivation domain (amino acids 1-74) deleted and replaced by  
CC the domain 325-393 of p53. The present sequence is that of  
CC a specifically claimed p53 variant designated 393-325 and comprising  
CC the 325-393 domain, amino acids 75-325 of human wild-type p53 and a  
CC leucine zipper domain at the C-terminal. The p53 variants are  
CC more active and more stable tumour suppressors and apoptosis-inducing  
CC agents than wild-type p53 and are active where the wild-type protein  
CC is not, i.e. they are not inactivated by dominant negative or oncogenic  
CC mutants, nor by other cellular proteins (because the leucine zipper  
CC domain prevents formation of inactive mixed oligomers).

SQ Sequence 353 AA;  
Query Match 100.0%; Score 58; DB 1; Length 353;  
Best Local Similarity 100.0%; Pred. No. 7.32e-01;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Db 38 GSRHSSHL 46  
QY 1 GSRHSSHL 9  
RESULT 10  
ID R51874 standard; Protein; 354 AA.  
AC R51874;  
DT 18-NOV-1994 (first entry)  
DE Human p53 amino acids 40-393.  
KW Human nuclear phosphoprotein p53; tumour suppressor gene product;  
KW anti-oncogene; cancer; tumour; antibody binding region; epitope.  
OS Homo sapiens.  
FH Key Location/Qualifiers  
FT Misc\_difference 234 /note= "Arg corresponds to a CAT codon"  
PN WO9408241-A.  
PD 14-APR-1994.  
PF 30-SEP-1993; E02666.  
PR 30-SEP-1992; DE-232823.  
PA (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM STIFTUNG.  
PI Klein R, Schranz P, Tessmer C, Volkman M, Zentgraf H;  
DR WPI: 94-135732/16.  
DR N-PSDB: Q62359.  
PT Non-radioactive detection of p53 specific antibodies - by capture  
PT on immobilised p53 or its fragments, then reaction with labelled  
PT second antibody, for diagnosis of tumours and suitable for  
PT screening  
PS Claim 10; Page 18; 35pp; German.  
CC Antibodies specific for p53 are detected by binding to immobilised  
CC fragments of the p53 gene product containing the antibody-binding  
CC region. Preferred fragments contain amino acids 1-241, 40-349,  
CC 40-393, 66-241, 66-393, 237-349, 237-393 and esp. 9-33, 37-52 or  
CC 368-386. See R51872-R51881 for sequences of these fragments.  
SQ Sequence 354 AA;  
Query Match 100.0%; Score 58; DB 1; Length 354;  
Best Local Similarity 100.0%; Pred. No. 7.32e-01;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Db 322 GSRHSSHL 330  
QY 1 GSRHSSHL 9  
RESULT 11  
ID Y03191 standard; Protein; 393 AA.  
AC Y03191;  
DT 21-JUN-1999 (first entry)  
DE Amino acid sequence of tumour suppressor p53.  
KW Human; p53 protein; mutant; tumour; colorectal cancer;  
KW transgenic animal.  
OS Homo sapiens.  
FH Key Location/Qualifiers  
FT Misc\_difference 224 /label= Glu  
FT /note= "encoded by GAGGAG"  
FT Misc\_difference 374 /note= "Gly  
FT /label= Gly  
FT /note= "encoded by GCT"  
PN WO9902682-A1.  
PD 21-JAN-1999.  
PF 09-JUL-1998; UI3949.  
PR 12-JAN-1998; WO-U01206.  
PR 09-JUL-1997; WO-052805.  
PA (AFFY-) AFFYMETRIX INC.  
PI Mack DH;

```

RESULT 5
ID W09322 standard; peptide; 74 AA.
AC W09322;
DT 10-JUN-1997 (first entry)
DE C-terminal domain of p53 protein.
KW Chimaeric; bispecific; DNA binding domain; trans; activator; repressor;
KW diphtheria; Pseudomonas; toxin; thymidine kinase; single chain antibody;
KW pathogen; HIV Tat; papilloma virus; E6/E7; Epstein-Barr virus; EBNA;
KW hyperproliferation; p53; tumour; oligomerisation.
OS Homo sapiens.
PN W0930512-A1.
PD 03-OCT-1996.
PF 29-MAR-1996; F00477.
PR 31-MAR-1995; FR-003841.
PA (RHON ) RHONE POULENC RORER SA.
PI Bracco L, Schweighoffer F, Tocque B;
DR WPI; 96-45359/45.
CC Conditional gene expression system triggered by e.g. infection or
PT hyper-proliferation - comprises novel bi-specific proteins having
PT DNA-binding domain and second domain specific for trans-activator or
PT repressor, for gene therapy
PS Claim 16; Page 44; 81pp; French.
CC The invention relates to novel chimaeric, bispecific proteins which
CC comprise: (a) a DNA binding domain and (b) a domain which binds a
CC trans-activator (TA), trans-repressor (TR) or their complexes, which are
CC characteristic of a physiological or physiopathological state. The novel
CC chimaeric, bispecific proteins allow expression of a therapeutic protein
CC (e.g. diphtheria or Pseudomonas toxins, thymidine kinase, single chain
CC antibodies) to be regulated in response to particular conditions.
CC Examples include making the protein responsive to the presence of
CC particular pathogenic TA mols (e.g. HIV Tat, papilloma virus E6/E7
CC proteins or Epstein-Barr virus EBNA protein), the therapeutic protein
CC will be expressed in those cells infected by that pathogen. Similarly,
CC where the chimaeric protein responds to a cellular protein typical of a
CC hyperproliferative state (esp. wild-type and mutant p53), expression can
CC be restricted to tumour cells. The sequence presented here is an example
CC of a TA binding domain. It corresponds to the C-terminal domain of the
CC p53 protein between residues 320-393 containing the oligomerisation
CC domain which binds TA proteins.
SQ Sequence 74 AA;

Query Match 100.0%; Score 58; DB 1; Length 74;
Best Local Similarity 100.0%; Pred. No. 7.32e-01;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 42 GSRHSSHL 50
QY 1 GSRHSSHL 9
|||||

RESULT 6
ID R51878 standard; Protein; 157 AA.
AC R51878;
DT 18-NOV-1994 (first entry)
DE Human p53 amino acids 237-393.
KW Human nuclear phosphoprotein p53; tumour suppressor gene product;
KW anti-oncogene; cancer; tumour; antibody binding region; epitope.
OS Homo sapiens.
FS Key Location/Qualifiers
FT misc_difference 37
FT misc_difference 37 /note= "Arg corresponds to a CAT codon"
PN W09408241-A.
PD 14-APR-1994.
PF 30-SEP-1993; E02666.
PR 30-SEP-1992; DE-232823.
PA (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM STIFTUNG.
PI Klein R, Schranz P, Tesser C, Volkmann M, Zentgraf H;
DR WPI; 94-135732/16.
DR N-P5DB; Q62363.
PT Non-radioactive detection of p53 specific antibodies - by capture
PT on immobilised p53 or its fragments, then reaction with labelled
PT second antibody, for diagnosis of tumours and suitable for
PT screening
PS Claim 10; Page 18; 35pp; German.
CC Antibodies specific for p53 are detected by binding to immobilised
CC fragments of the p53 gene product containing the antibody-binding
CC region. Preferred fragments contain amino acids 1-241, 40-349,
CC 40-393, 66-241, 66-393, 237-349, 237-393 and esp. 9-33, 37-52 or
CC 368-386. See R51872-R51881 for sequences of these fragments.
SQ Sequence 157 AA;

Query Match 100.0%; Score 58; DB 1; Length 157;
Best Local Similarity 100.0%; Pred. No. 7.32e-01;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 125 GSRHSSHL 133
QY 1 GSRHSSHL 9
|||||

RESULT 7
ID R51876 standard; Protein; 328 AA.
AC R51876;
DT 18-NOV-1994 (first entry)
DE Human p53 amino acids 66-393.
KW Human nuclear phosphoprotein p53; tumour suppressor gene product;
KW anti-oncogene; cancer; tumour; antibody binding region; epitope.
OS Homo sapiens.
FS Key Location/Qualifiers
FT misc_difference 208
FT misc_difference 208 /note= "Arg corresponds to a CAT codon"
PN W09408241-A.
PD 14-APR-1994.
PF 30-SEP-1993; E02666.
PR 30-SEP-1992; DE-232823.
PA (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM STIFTUNG.
PI Klein R, Schranz P, Tesser C, Volkmann M, Zentgraf H;
DR WPI; 94-135732/16.
DR N-P5DB; Q62361.
PT Non-radioactive detection of p53 specific antibodies - by capture
PT on immobilised p53 or its fragments, then reaction with labelled
PT second antibody, for diagnosis of tumours and suitable for
PT screening
PS Claim 10; Page 18; 35pp; German.
CC Antibodies specific for p53 are detected by binding to immobilised
CC fragments of the p53 gene product containing the antibody-binding
CC region. Preferred fragments contain amino acids 1-241, 40-349,
CC 40-393, 66-241, 66-393, 237-349, 237-393 and esp. 9-33, 37-52 or
CC 368-386. See R51872-R51881 for sequences of these fragments.
SQ Sequence 328 AA;

Query Match 100.0%; Score 58; DB 1; Length 328;
Best Local Similarity 100.0%; Pred. No. 7.32e-01;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 296 GSRHSSHL 304
QY 1 GSRHSSHL 9
|||||

RESULT 8
ID W28494 standard; Protein; 353 AA.
AC W28494;
DT 25-NOV-1997 (first entry)
DE Human p53 protein variant 393-325H.
KW Leucine zipper domain; LZD; oligomerisation domain; mutant; mutein;
KW substitution; replacement; transactivation; viral protein VP16; HSV;
KW anti-oncogene; hyperproliferation; cancer; restenosis;
KW tumour suppression; apoptosis.
OS Homo sapiens.
FS Key Location/Qualifiers
FT misc_difference 179
FT misc_difference 179 /note= "Arg residue at position 182 of wild-type
FT on immobilised p53 or its fragments, then reaction with labelled
FT second antibody, for diagnosis of tumours and suitable for
PT screening
PN W09704092-A1.

```

PD 16-APR-1998.  
PF 01-OCT-1997; U16132.  
PR 07-OCT-1996; US-028533.  
PA (SCHE ) SCHERING CORP.  
PI Mytych DT, Swanson SV;  
DR WPI: 98-240965/21.  
PT Detecting antibodies that bind p53 by reaction with immobilised p53  
PT peptide(s) - attached directly to flow cells in the sensor chip of  
PT bio-sensor, used to analyse serum from cancer patients, e.g. those  
PT being given p53 gene therapy  
PS Claim 4; Page 5; 41pp; English.  
CC Peptides W60202-05 are derived from human p53 protein. The present  
CC peptide corresponds to residues 346-370. The peptides are used in  
CC the method of the invention. Antibodies that bind to p53 protein are  
CC detected by immobilising a p53 peptide directly on to a flow cell of  
CC a sensor chip in a biosensor, treating the peptide with a sample of  
CC patient serum, diluted in buffer, and measuring binding of antibody  
CC to the peptide using the biosensor. The method is used to monitor  
CC cancer patients undergoing p53 gene therapy (to determine if an  
CC immune response has developed), and also to detect antibodies against  
CC mutant forms of p53.  
SQ Sequence 26 AA;

Query Match 100.0%; Score 58; DB 1; Length 26;  
Best Local Similarity 100.0%; Pred. No. 7.32e-01;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 17 GSRHSSHL 25  
QY 1 GSRHSSHL 9

RESULT 3  
ID W05368 standard; peptide: 26 AA.  
AC W05368;  
DT 30-APR-1997 (first entry)  
DE Peptide p53p360-386DGF.  
KW Human; p53; cell proliferation; cell death; regulator; tumour; psoriasis;  
KW negative regulatory region; DNA damaging agent; transplant rejection;  
KW abnormal cell proliferation; atherosclerosis; cancer; autoimmune disease;  
KW arterial restenosis; immune response; apoptosis; inducer; therapy;  
KW proliferating lymphocytes.  
OS Synthetic.  
PN W09625434-A1.  
PD 22-AUG-1996.  
PF 16-FEB-1996; U01535.  
PR 16-FEB-1995; US-392542.  
PA (FARB ) BAYER CORP.  
PI Halazonetis T, Hartwig W;  
DR WPI: 96-393345/39.  
PT New human p53-isomorph peptide(s) and peptidomimetic cpds. - used  
PT for activating p53 function, e.g. for treating tumours, cancers,  
PT psoriasis, etc  
PS Disclosure: Page 12; 55pp; English.  
CC W05365-W05374 represent examples of the p53 (see W05344 for full length  
CC wild type sequence) peptides of the invention. These sequences all have  
CC additions or deletions of residues from the wild type peptide fragments  
CC of the invention (see W05350-W05364). The p53 protein functions to  
CC regulate cell proliferation and cell death, and is mutated in more than  
CC half of all human tumours. These sequences are used to activate the DNA  
CC binding activity of wild type p53, and p53 mutants (see W05345-W05349).  
CC The peptides of the invention consist of at least four sequential amino  
CC acids from a negative regulatory region which maps to residues 361-383 of  
CC p53. These sequences preferably contain four amino acids from a non-human  
CC p53 sequence, contain D-form amino acids, and can also be cyclic  
CC peptides. The sequences retain the structural characteristics of the  
CC original peptides, but the modifications render them less susceptible to  
CC cleavage by proteases and exopeptidases. As these sequences activate p53  
CC DNA binding, they can be used to identify p53 mutants. The peptides can  
CC also be used for treating a patient with a tumour expressing a p53 mutant  
CC whose ability to bind DNA may be activated by one of the peptides. They  
CC can also be used for treating conditions such as exposure to DNA damaging

CC agents, abnormal cell proliferation characteristic of psoriasis,  
CC atherosclerosis, cancer, arterial restenosis, autoimmune diseases and  
CC undesirable immune responses accompanying rejection of a transplant. The  
CC peptides can also induce apoptosis of specific cells, such as  
CC proliferating lymphocytes.  
SQ Sequence 26 AA;

Query Match 100.0%; Score 58; DB 1; Length 26;  
Best Local Similarity 100.0%; Pred. No. 7.32e-01;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 2 GSRHSSHL 10  
QY 1 GSRHSSHL 9

RESULT 4  
ID W05366 standard; peptide: 27 AA.  
AC W05366;  
DT 30-APR-1997 (first entry)  
DE Peptide p53p360-386DGF.  
KW Human; p53; cell proliferation; cell death; regulator; tumour; psoriasis;  
KW negative regulatory region; DNA damaging agent; transplant rejection;  
KW abnormal cell proliferation; atherosclerosis; cancer; autoimmune disease;  
KW arterial restenosis; immune response; apoptosis; inducer; therapy;  
KW proliferating lymphocytes.  
OS Synthetic.  
PN W09625434-A1.  
PD 22-AUG-1996.  
PF 16-FEB-1996; U01535.  
PR 16-FEB-1995; US-392542.  
PA (FARB ) BAYER CORP.  
PI Halazonetis T, Hartwig W;  
DR WPI: 96-393345/39.  
PT New human p53-isomorph peptide(s) and peptidomimetic cpds. - used  
PT for activating p53 function, e.g. for treating tumours, cancers,  
PT psoriasis, etc  
PS Disclosure: Page 12; 55pp; English.  
CC W05365-W05374 represent examples of the p53 (see W05344 for full length  
CC wild type sequence) peptides of the invention. These sequences all have  
CC additions or deletions of residues from the wild type peptide fragments  
CC of the invention (see W05350-W05364). The p53 protein functions to  
CC regulate cell proliferation and cell death, and is mutated in more than  
CC half of all human tumours. These sequences are used to activate the DNA  
CC binding activity of wild type p53, and p53 mutants (see W05345-W05349).  
CC The peptides of the invention consist of at least four sequential amino  
CC acids from a negative regulatory region which maps to residues 361-383 of  
CC p53. These sequences preferably contain four amino acids from a non-human  
CC p53 sequence, contain D-form amino acids, and can also be cyclic  
CC peptides. The sequences retain the structural characteristics of the  
CC original peptides, but the modifications render them less susceptible to  
CC cleavage by proteases and exopeptidases. As these sequences activate p53  
CC DNA binding, they can be used to identify p53 mutants. The peptides can  
CC also be used for treating a patient with a tumour expressing a p53 mutant  
CC whose ability to bind DNA may be activated by one of the peptides. They  
CC can also be used for treating conditions such as exposure to DNA damaging

Query Match 100.0%; Score 58; DB 1; Length 27;  
Best Local Similarity 100.0%; Pred. No. 7.32e-01;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 2 GSRHSSHL 10  
QY 1 GSRHSSHL 9



\*\*\*\*\*  
WIP  
\*\*\*\*\*  
(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

Mpsrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Sat Apr 15 01:11:47 2000; MasPar time 3.11 Seconds  
Tabular output not generated. 68.479 Million cell updates/sec

Title: >US-08-452-843-21  
Description: (1-9) from US08452843.pap  
Perfect Score: 58  
Sequence: 1 GSRAHSSHL 9  
Scoring table: PAM 150  
Gap 15  
Searched: 188963 seqs, 23686106 residues  
Post-processing: Minimum Match 0%  
Listing first 45 summaries  
Database: a-geneseq36  
1:geneseq  
Statistics: Mean 14.751; Variance 35.861; scale 0.411

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	58	100.0	15	1 R54917	Immunodominant epitope	7.32e-01
2	58	100.0	26	1 W60204	p53 peptide used to de	7.32e-01
3	58	100.0	26	1 W05368	Peptide p53PC360-386DG	7.32e-01
4	58	100.0	27	1 W05366	Peptide p53PC360-386DG	7.32e-01
5	58	100.0	74	1 W09322	C-terminal domain of p	7.32e-01
6	58	100.0	157	1 R51878	Human p53 amino acids	7.32e-01
7	58	100.0	328	1 R51876	Human p53 amino acids	7.32e-01
8	58	100.0	353	1 W28494	Human p53 protein vari	7.32e-01
9	58	100.0	353	1 W28493	Human p53 protein vari	7.32e-01
10	58	100.0	354	1 R51874	Human p53 amino acids	7.32e-01
11	58	100.0	393	1 Y03191	Amino acid sequence of	7.32e-01
12	58	100.0	393	1 W02617	Human p53 tumour suppr	7.32e-01
13	58	100.0	393	1 W84270	Human p53 protein.	7.32e-01
14	58	100.0	393	1 W69219	Human p53 mutant 2.	7.32e-01
15	58	100.0	393	1 W69718	Human p53 used in coup	7.32e-01
16	58	100.0	393	1 W05348	Human p53 mutant R282W	7.32e-01
17	58	100.0	393	1 W05344	Human p53.	7.32e-01
18	58	100.0	393	1 W07243	Human p53 protein SEQ	7.32e-01
19	58	100.0	393	1 W13970	Modified p53 variant p	7.32e-01
20	58	100.0	393	1 W57242	Human p53 protein SEQ	7.32e-01
21	58	100.0	393	1 W48658	Amino acid sequence of	7.32e-01
22	58	100.0	393	1 W05346	Human p53 mutant R273H	7.32e-01
23	58	100.0	393	1 W05347	Human p53 mutant R248Q	7.32e-01

24	58	100.0	393	1 W69217	Human wild-type p53 pr	7.32e-01
25	58	100.0	393	1 W69218	Human p53 mutant 1.	7.32e-01
26	58	100.0	393	1 W13968	Modified p53 variant p	7.32e-01
27	58	100.0	393	1 R26758	p53.	7.32e-01
28	58	100.0	393	1 W13953	T284K modified human p	7.32e-01
29	58	100.0	393	1 W13980	Human tumour-derived p	7.32e-01
30	58	100.0	393	1 W05345	Human p53 mutant N239S	7.32e-01
31	58	100.0	393	1 W57244	Human p53 protein SEQ	7.32e-01
32	58	100.0	393	1 W13981	Human tumour-derived p	7.32e-01
33	58	100.0	393	1 R94623	p53 protein.	7.32e-01
34	58	100.0	393	1 R22238	Sequence of 53 kD cell	7.32e-01
35	58	100.0	393	1 W57245	Human p53 protein SEQ	7.32e-01
36	58	100.0	393	1 W13979	Human tumour-derived p	7.32e-01
37	58	100.0	401	1 W28487	Human p53 protein vari	7.32e-01
38	58	100.0	402	1 W13965	Chimeric p53 protein.	7.32e-01
39	58	100.0	404	1 W13963	Chimeric p53 protein.	7.32e-01
40	58	100.0	406	1 W13964	Chimeric p53 protein.	7.32e-01
41	58	100.0	406	1 W13966	Chimeric p53 protein.	7.32e-01
42	58	100.0	411	1 W13967	Chimeric p53 protein.	7.32e-01
43	58	100.0	438	1 R74272	tumour suppressor prot	7.32e-01
44	58	100.0	438	1 R50088	p53 tumour suppressor	7.32e-01
45	58	100.0	533	1 W19763	p53-GM-CSF immunostimu	7.32e-01

ALIGNMENTS

RESULT 1  
ID R54917 standard; peptide; 15 AA.  
AC R54917;  
DT 29-NOV-1994 (first entry)  
DE Immunodominant epitope from p53 C-terminal.  
KW cancer; pre-cancerous state; detection; diagnosis; human p53 gene;  
KW immunodominant epitope; human cellular tumour antigen;  
KW transformation-associated protein.  
OS Homo sapiens.  
PN W09410306-A.  
PD 11-MAY-1994.  
PF 02-NOV-1993; F01082.  
PR 02-NOV-1992; FR-013110.  
PA (EURO-) LAB EURO BIO SA.  
PI Legros Y, Lubin R, Soussi T;  
DR WFI; 94-167463/20.  
PT New immuno-dominant epitope(s) of protein p53 - for detecting and  
PT monitoring antibodies indicative of cancer and precancerous  
PT states  
PS Claim 7; Page 43; 62pp; French.  
CC Peptides derived from the N-terminal (amino acids 1-112) or the C-  
CC terminal (amino acids 350-393) of protein p53 which specifically  
CC react with anti-p53 antibodies in patients with cancer or  
CC precancerous conditions are claimed. The peptides (R54907-R54921)  
CC are useful for detecting and monitoring cancerous and precancerous  
CC conditions.  
SQ Sequence 15 AA;

Query Match 100.0%; Score 58; DB 1; Length 15;  
Best Local Similarity 100.0%; Pred. No. 7.32e-01;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 GSRAHSSHL 9  
| | | | | | | | | |  
Qy 1 GSRAHSSHL 9

RESULT 2  
ID W60204 standard; peptide; 26 AA.  
AC W60204;  
DT 18-AUG-1998 (first entry)  
DE p53 peptide used to detect antibodies against p53.  
KW Human; p53; antibody; detection; biosensor; cancer patient;  
KW p53 gene therapy; immune response; mutant.  
OS Synthetic.  
OS Homo sapiens.  
PN W09815834-A1.



PAGE BLANK (USPTO)

```
CC      CYCLIN-DEPENDENT KINASES (BY SIMILARITY).
CC      -!- SUBCELLULAR LOCATION: NUCLEAR.
DR      EMBL; X60018; CAA42633.1; -.
DR      HSSP; P04637; 1SAH.
DR      PROSITE; PS00348; P53; 1.
DR      PFAM; PF00870; P53; 1.
KW      Anti-oncogene; DNA-binding; Transcription regulation; Activator;
KW      Nuclear protein; Phosphorylation.
FT      VARIANT 163 163 H -> Y.
FT      NON_TER 393 393
SQ      SEQUENCE 393 AA; 43627 MW; AFD8A9E3 CRC32;

Query Match      100.0%; Score 78; DB 4; Length 393;
Best Local Similarity 100.0%; Pred. No. 3.36e-05;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 321 KPLDGEYFTL 330
QY 1 KPLDGEYFTL 10

RESULT 13
ID Q16535 PRELIMINARY; PRT; 393 AA.
AC Q16535;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DE P53 TRANSFORMATION SUPPRESSOR (FRAGMENT).
GN P53.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 92007731.
RA FARRELL P.J., ALLAN G., SHANAHAN F., VOUSDEN K.H., CROOK T.;
RT "P53 is frequently mutated in Burkitt's lymphoma cell lines.";
RL EMBL J. 10:2879-2887(1991).
DR      EMBL; X60017; CAA42632.1; -.
DR      EMBL; X60015; CAA42630.1; -.
DR      HSSP; P04637; 1SAH.
DR      PFAM; PF00870; P53; 1.
FT      VARIANT 248 248 Q -> R.
FT      NON_TER 393 393
SQ      SEQUENCE 393 AA; 43684 MW; 239818A9 CRC32;

Query Match      100.0%; Score 78; DB 4; Length 393;
Best Local Similarity 100.0%; Pred. No. 3.36e-05;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 321 KPLDGEYFTL 330
QY 1 KPLDGEYFTL 10

RESULT 14
ID Q9WUR6 PRELIMINARY; PRT; 391 AA.
AC Q9WUR6;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DE 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
DE CELLULAR TUMOR ANTIGEN P53.
GN P53.
OS Cavia porcellus (Guinea pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Hystricognathi; Caviidae; Cavia.
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE-SPLEEN;
RC MEDLINE; 99265972.
RA D'ERCHIA A.M., PESOLE G., TULLIO A., SACCONI C., SBISA E.;
RT "Guinea pig p53 mRNA: identification of new elements in coding and
RT - untranslated regions and their functional and evolutionary
```

```
RT      Implications.";
RL      Genomics 58:50-64(1999).
CC      -!- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT
CC      PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL
CC      CYCLE REGULATION, IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY
CC      REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED
CC      FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF
CC      CYCLIN-DEPENDENT KINASES (BY SIMILARITY).
CC      -!- SUBCELLULAR LOCATION: NUCLEAR.
DR      EMBL; AJ009673; CAB43196.1; -.
DR      PROSITE; PS00348; P53; 1.
KW      Anti-oncogene; DNA-binding; Transcription regulation; Activator;
KW      Nuclear protein; Phosphorylation.
SQ      SEQUENCE 391 AA; 43288 MW; BFD34AB4 CRC32;

Query Match      92.3%; Score 72; DB 11; Length 391;
Best Local Similarity 90.0%; Pred. No. 8.86e-04;
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 319 KPLDAEYFTL 328
QY 1 KPLDGEYFTL 10

RESULT 15
ID Q29484 PRELIMINARY; PRT; 196 AA.
AC Q29484;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
DE CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).
GN P53.
OS Equus caballus (Horse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Perissodactyla; Equidae; Equus.
RN [1]
RP SEQUENCE FROM N.A.
RA BUCHER K., SZALAI G., MARTI E., PAULI U., LAZARY S.;
RL Res. Vet. Sci. 0:0-0(0).
CC      -!- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT
CC      PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL
CC      CYCLE REGULATION, IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY
CC      REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED
CC      FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF
CC      CYCLIN-DEPENDENT KINASES (BY SIMILARITY).
CC      -!- SUBCELLULAR LOCATION: NUCLEAR.
DR      EMBL; X91793; CAA62905.1; -.
DR      HSSP; P04637; 1SAH.
DR      PROSITE; PS00348; P53; 1.
DR      PFAM; PF00870; P53; 1.
KW      Anti-oncogene; DNA-binding; Transcription regulation; Activator;
KW      Nuclear protein; Phosphorylation.
FT      NON_TER 196 196
FT      NON_TER 196 196
SQ      SEQUENCE 196 AA; 22080 MW; F443239C CRC32;

Query Match      89.7%; Score 70; DB 6; Length 196;
Best Local Similarity 100.0%; Pred. No. 2.57e-03;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 188 KPLDGEYFTL 196
QY 1 KPLDGEYFTL 9

Search completed: Sat Apr 15 01:08:24 2000
Job time : 93 secs.
```

DR PFAM; PF00870; P53; 1.  
KW Repeat; Tumor antigen; Anti-oncogene; DNA-binding;  
FT NON\_TER 393  
SQ SEQUENCE 393 AA; 43698 MW; 3EA71431 CRC32;

Query Match 100.0%; Score 78; DB 4; Length 393;  
Best Local Similarity 100.0%; Pred. No. 3.36e-05;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 321 KPLDGEYFTL 330  
QY 1 KPLDGEYFTL 10

RESULT 9  
ID Q15087 PRELIMINARY; PRT; 393 AA.  
AC Q15087;  
DT 01-NOV-1996 (TREMBlrel. 01, Created)  
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
DE P53 TRANSFORMATION SUPPRESSOR (FRAGMENT).  
GN P53.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 92007731.  
RA FARRELL P.J., ALLAN G., SHANAHAN F., VOUSDEN K.H., CROOK T.;  
RT "p53 is frequently mutated in Burkitt's lymphoma cell lines."  
RL EMO J. 10:2879-2887(1991).  
DR EMBL; X60014; CAA42629.1; -.  
DR HSSP; P04637; 1SAH.  
DR PFAM; PF00870; P53; 1.  
FT VARIANT 237 237 I -> M.  
FT NON\_TER 393 393  
SQ SEQUENCE 393 AA; 43694 MW; 9B881992 CRC32;

Query Match 100.0%; Score 78; DB 4; Length 393;  
Best Local Similarity 100.0%; Pred. No. 3.36e-05;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 321 KPLDGEYFTL 330  
QY 1 KPLDGEYFTL 10

RESULT 10  
ID Q15088 PRELIMINARY; PRT; 393 AA.  
AC Q15088;  
DT 01-NOV-1996 (TREMBlrel. 01, Created)  
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
DE P53 TRANSFORMATION SUPPRESSOR (FRAGMENT).  
GN P53.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 92007731.  
RA FARRELL P.J., ALLAN G., SHANAHAN F., VOUSDEN K.H., CROOK T.;  
RT "p53 is frequently mutated in Burkitt's lymphoma cell lines."  
RL EMO J. 10:2879-2887(1991).  
DR EMBL; X60016; CAA42631.1; -.  
DR HSSP; P04637; 1SAH.  
DR PFAM; PF00870; P53; 1.  
FT VARIANT 238 238 Y -> C.  
FT NON\_TER 393 393  
SQ SEQUENCE 393 AA; 43713 MW; A01E1523 CRC32;

Query Match 100.0%; Score 78; DB 4; Length 393;

Best Local Similarity 100.0%; Pred. No. 3.36e-05;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Db 321 KPLDGEYFTL 330  
QY 1 KPLDGEYFTL 10

RESULT 11  
ID Q16848 PRELIMINARY; PRT; 393 AA.  
AC Q16848;  
DT 01-NOV-1996 (TREMBlrel. 01, Created)  
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
DE CELLULAR TUMOR ANTIGEN P53.  
GN TP53.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 87089826.  
RA HARRIS N., BRILL E., SHOHAT O., PROKOCIMER M., WOLF D., ARAI N.,  
RA ROTHER V.;  
RT "Molecular basis for heterogeneity of the human p53 protein."  
RL Mol. Cell. Biol. 6:4650-4656(1986).  
CC -1- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT  
CC PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL  
CC CYCLE REGULATION, IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY  
CC REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED  
CC FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF  
CC CYCLIN-DEPENDENT KINASES (BY SIMILARITY).  
CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
DR EMBL; M14694; AAA61211.1; -.  
DR HSSP; P04637; 1TSR.  
DR PROSITE; PS00348; P53; 1.  
DR PFAM; PF00870; P53; 1.  
DR PRINTS; PR00386; P53SUPPRESSR.  
KW Nuclear protein; Phosphorylation; Anti-oncogene; DNA-binding;  
KW Transcription regulation; Activator.  
SQ SEQUENCE 393 AA; 43723 MW; DA7D302F CRC32;

Query Match 100.0%; Score 78; DB 4; Length 393;  
Best Local Similarity 100.0%; Pred. No. 3.36e-05;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 321 KPLDGEYFTL 330  
QY 1 KPLDGEYFTL 10

RESULT 12  
ID Q16808 PRELIMINARY; PRT; 393 AA.  
AC Q16808;  
DT 01-NOV-1996 (TREMBlrel. 01, Created)  
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
DE CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).  
GN P53.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 92007731.  
RA FARRELL P.J., ALLAN G., SHANAHAN F., VOUSDEN K.H., CROOK T.;  
RT "p53 is frequently mutated in Burkitt's lymphoma cell lines."  
RL EMO J. 10:2879-2887(1991).  
CC -1- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT  
CC PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL  
CC CYCLE REGULATION, IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY  
CC REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED  
CC FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF

CC -1- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT  
 CC PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL  
 CC CYCLE REGULATION, IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY  
 CC REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED  
 CC FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF  
 CC CYCLIN-DEPENDENT KINASES (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 DR EMBL; X60011; CAA42626.1; -.  
 DR HSSP; P04637; 1SAH.  
 DR PFAM; PF00870; P53; 1.  
 DR PROSITE; PS00348; P53; 1.  
 KW Anti-oncogene; DNA-binding; Transcription regulation; Activator;  
 KW Nuclear protein; Phosphorylation.  
 FT VARIANT 193 193 R -> H.  
 FT NON\_TER 393 393  
 SQ SEQUENCE 393 AA; 43731 MW; 279BC9CB CRC32;

Query Match 100.0%; Score 78; DB 4; Length 393;  
 Best Local Similarity 100.0%; Pred. No. 3.36e-05;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 321 KPLDGEYFTL 330  
 QY 1 KPLDGEYFTL 10  
 RESULT 6  
 ID Q15086 PRELIMINARY; PRT; 393 AA.  
 AC Q15086;  
 DT 01-NOV-1996 (TRENBLrel. 01, Created)  
 DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)  
 DT 01-NOV-1999 (TRENBLrel. 12, Last annotation update)  
 DE P53 TRANSFORMATION SUPPRESSOR (FRAGMENT).  
 GN P53.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 92007731.  
 RA FARRELL P.J., ALLAN G., SHANAHAN F., VOUSDEN K.H., CROOK T.;  
 RT "P53 is frequently mutated in Burkitt's lymphoma cell lines."  
 RL EMBO J. 10:2879-2887(1991).  
 DR EMBL; X60013; CAA42628.1; -.  
 DR HSSP; P04637; 1SAH.  
 DR PFAM; PF00870; P53; 1.  
 DR VARIANT 246 246 T -> M.  
 FT NON\_TER 393 393  
 SQ SEQUENCE 393 AA; 43682 MW; 943B62A3 CRC32;

Query Match 100.0%; Score 78; DB 4; Length 393;  
 Best Local Similarity 100.0%; Pred. No. 3.36e-05;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 321 KPLDGEYFTL 330  
 QY 1 KPLDGEYFTL 10  
 RESULT 7  
 ID Q16810 PRELIMINARY; PRT; 393 AA.  
 AC Q16810;  
 DT 01-NOV-1996 (TRENBLrel. 01, Created)  
 DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)  
 DT 01-NOV-1999 (TRENBLrel. 12, Last annotation update)  
 DE CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).  
 GN P53.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 92007731.

RA FARRELL P.J., ALLAN G., SHANAHAN F., VOUSDEN K.H., CROOK T.;  
 RT "P53 is frequently mutated in Burkitt's lymphoma cell lines."  
 RL EMBO J. 10:2879-2887(1991).  
 CC -1- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT  
 CC PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL  
 CC CYCLE REGULATION, IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY  
 CC REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED  
 CC FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF  
 CC CYCLIN-DEPENDENT KINASES (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 DR EMBL; X60020; CAA42635.1; -.  
 DR HSSP; P04637; 1SAH.  
 DR PROSITE; PS00348; P53; 1.  
 DR PFAM; PF00870; P53; 1.  
 KW Anti-oncogene; DNA-binding; Transcription regulation; Activator;  
 KW Nuclear protein; Phosphorylation.  
 FT VARIANT 254 254 D -> N.  
 FT VARIANT 254 254 D -> V.  
 FT NON\_TER 393 393  
 SQ SEQUENCE 393 AA; 43714 MW; 5F914579 CRC32;

Query Match 100.0%; Score 78; DB 4; Length 393;  
 Best Local Similarity 100.0%; Pred. No. 3.36e-05;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 321 KPLDGEYFTL 330  
 QY 1 KPLDGEYFTL 10  
 RESULT 8  
 ID Q16811 PRELIMINARY; PRT; 393 AA.  
 AC Q16811;  
 DT 01-NOV-1996 (TRENBLrel. 01, Created)  
 DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)  
 DT 01-NOV-1999 (TRENBLrel. 12, Last annotation update)  
 DE CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 85126934.  
 RA MATLASHESKI G., LAMB P., PIM D., PEACOCK J., CRAWFORD L.,  
 RA BENCHIMOL S.;  
 RT "Isolation and characterization of a human p53 cDNA clone: expression  
 RT of the human p53 gene."  
 RL EMBO J. 3:3257-3262(1984).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 87064416.  
 RA LAMB P., CRAWFORD L.;  
 RT "Characterization of the human p53 gene."  
 RL Mol. Cell. Biol. 6:1379-1385(1986).  
 CC -1- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT  
 CC PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL  
 CC CYCLE REGULATION, IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY  
 CC REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED  
 CC FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF  
 CC CYCLIN-DEPENDENT KINASES (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 DR EMBL; M13121; AAA59987.1; -.  
 DR EMBL; M13112; AAA59987.1; JOINED.  
 DR EMBL; M13113; AAA59987.1; JOINED.  
 DR EMBL; M13114; AAA59987.1; JOINED.  
 DR EMBL; M13115; AAA59987.1; JOINED.  
 DR EMBL; M13116; AAA59987.1; JOINED.  
 DR EMBL; M13117; AAA59987.1; JOINED.  
 DR EMBL; M13118; AAA59987.1; JOINED.  
 DR EMBL; M13119; AAA59987.1; JOINED.  
 DR EMBL; M13120; AAA59987.1; JOINED.  
 DR HSSP; P04637; ITSK.  
 DR PROSITE; PS00348; P53; 1.

```
RESULT 2
ID O70366 PRELIMINARY; PRT; 390 AA.
AC O70366;
DT 01-AUG-1998 (TREMBlrel. 07, Created)
DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)
DE 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
DE CELLULAR TUMOR ANTIGEN P53.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-LYMPHOID LEUKEMIA;
RA FROSTESJO L., NILSSON J., WANDZIOCH E., HEBY O.;
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT
CC PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL
CC CYCLE REGULATION, IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY
CC REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED
CC FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF
CC CYCLIN-DEPENDENT KINASES (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
DR EMBL; AF051368; AAC05704.1; -.
DR HSSP; P04637; IPET.
DR PROSITE; PS00348; P53; 1.
DR PFAM; PF00870; P53; 1.
DR PRINTS; PR00386; P53SUPPRESSR.
KW Anti-oncogene; DNA-binding; Transcription regulation; Activator;
KW Nuclear protein; Phosphorylation.
SQ SEQUENCE 390 AA; 43430 MW; EBF4C8AA CRC32;

Query Match 100.0%; Score 78; DB 11; Length 390;
Best Local Similarity 100.0%; Pred. No. 3.36e-05;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 318 KPLDGEYFTL 327
QY 1 KPLDGEYFTL 10
|||||

RESULT 3
ID O36006 PRELIMINARY; PRT; 391 AA.
AC O36006;
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DE 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
DE CELLULAR TUMOR ANTIGEN P53.
GN P53.
OS Marmota monax (Woodchuck).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Sciuridae; Sciurinae; Marmota.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 97376996.
RA FEITELSON M.A., RANGANATHAN P.N., CLAYTON M.M., ZHANG S.M.;
RT "Partial characterization of the woodchuck tumor suppressor, p53, and
RT its interaction with woodchuck hepatitis virus X antigen in
RT hepatocarcinogenesis."
RL Oncogene 15:327-336(1997).
CC -!- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT
CC PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL
CC CYCLE REGULATION, IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY
CC REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED
CC FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF
CC CYCLIN-DEPENDENT KINASES (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
DR EMBL; AJ001022; CAA04478.1; -.
DR HSSP; P04637; ITRR.
DR PROSITE; PS00348; P53; 1.
DR PFAM; PF00870; P53; 1.
DR PRINTS; PR00386; P53SUPPRESSR.

KW Anti-oncogene; DNA-binding; Transcription regulation; Activator;
KW Nuclear protein; Phosphorylation.
SQ SEQUENCE 390 AA; 43430 MW; EBF4C8AA CRC32;

Query Match 100.0%; Score 78; DB 11; Length 390;
Best Local Similarity 100.0%; Pred. No. 3.36e-05;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 318 KPLDGEYFTL 327
QY 1 KPLDGEYFTL 10
|||||

RESULT 4
ID Q16809 PRELIMINARY; PRT; 393 AA.
AC Q16809;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DE 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
DE CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).
GN P53.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Homnidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 92007731.
RA FARRELL P.J., ALLAN G., SHANAHAN F., VOUSDEN K.H., CROOK T.;
RT "p53 is frequently mutated in Burkitt's lymphoma cell lines."
RL EMBO J. 10:2879-2887(1991).
CC -!- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT
CC PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL
CC CYCLE REGULATION, IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY
CC REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED
CC FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF
CC CYCLIN-DEPENDENT KINASES (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
DR EMBL; X60019; CAA42634.1; -.
DR HSSP; P04637; ISAH.
DR PROSITE; PS00348; P53; 1.
DR PFAM; PF00870; P53; 1.
KW Anti-oncogene; DNA-binding; Transcription regulation; Activator;
KW Nuclear protein; Phosphorylation.
FT VARIANT 213 213 Q -> R.
FT NON_TER 393
SQ SEQUENCE 393 AA; 43684 MW; CB70BD7F CRC32;

Query Match 100.0%; Score 78; DB 4; Length 393;
Best Local Similarity 100.0%; Pred. No. 3.36e-05;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 321 KPLDGEYFTL 330
QY 1 KPLDGEYFTL 10
|||||

RESULT 5
ID Q16807 PRELIMINARY; PRT; 393 AA.
AC Q16807;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DE 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
DE CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).
GN P53.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Homnidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 92007731.
RA FARRELL P.J., ALLAN G., SHANAHAN F., VOUSDEN K.H., CROOK T.;
RT "p53 is frequently mutated in Burkitt's lymphoma cell lines."
RL EMBO J. 10:2879-2887(1991).
```

(TM)

Result No.	Query			ID	Description	Pred. No.	
	Score	Match	Length				
1	78	100.0	281	6	Q29475	CELLULAR TUMOR ANTIGEN	3.36e-05
2	78	100.0	390	11	Q70366	CELLULAR TUMOR ANTIGEN	3.36e-05
3	78	100.0	391	6	Q56006	CELLULAR TUMOR ANTIGEN	3.36e-05
4	78	100.0	393	4	Q16809	CELLULAR TUMOR ANTIGEN	3.36e-05
5	78	100.0	393	4	Q16807	CELLULAR TUMOR ANTIGEN	3.36e-05
6	78	100.0	393	4	Q15086	P53 TRANSFORMATION SUP	3.36e-05
7	78	100.0	393	4	Q16810	CELLULAR TUMOR ANTIGEN	3.36e-05
8	78	100.0	393	4	Q16811	CELLULAR TUMOR ANTIGEN	3.36e-05
9	78	100.0	393	4	Q15087	P53 TRANSFORMATION SUP	3.36e-05
10	78	100.0	393	4	Q15088	P53 TRANSFORMATION SUP	3.36e-05
11	78	100.0	393	4	Q16848	CELLULAR TUMOR ANTIGEN	3.36e-05
12	78	100.0	393	4	Q16808	CELLULAR TUMOR ANTIGEN	3.36e-05
13	78	100.0	393	4	Q16535	P53 TRANSFORMATION SUP	3.36e-05
14	72	92.3	391	11	Q5WUR6	CELLULAR TUMOR ANTIGEN	8.86e-04
15	70	89.7	196	6	Q29484	CELLULAR TUMOR ANTIGEN	2.57e-03
16	70	89.7	205	11	Q35873	CELLULAR TUMOR ANTIGEN	2.57e-03
17	69	88.5	238	14	P89004	P53 (FRAGMENT).	4.35e-03
18	69	88.5	286	14	P90332	P53 (FRAGMENT).	4.35e-03
19	69	88.5	286	14	P89003	P53 (FRAGMENT).	4.35e-03
20	69	88.5	378	14	P89002	P53 (FRAGMENT).	4.35e-03

—

```

Query Match      100.0%;      Score 78;  DB 6;  Length 281;
Best Local Similarity 100.0%;  Pred. No. 3.36e-05;
Matches 10;  Conservative 0;  Mismatches 0;  Indels
0      214  KPLDGEYFTL 223
          ||| |||||
          1  KPLDGEYFTL 10

```

```
CC BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2
CC EXPRESSION.
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY
CC OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED
CC IN MANY TYPES OF CANCER.
CC -!- SIMILARITY: BELONGS TO THE P53 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL; Y08900; CAA70108.1; -
CC DR EMBL; Y08901; CAA70109.1; -
CC DR EMBL; U50395; AAC53040.1; -
CC DR EMBL; D86070; BAAL3004.1; -
CC DR HSP; P04637; LYCO.
CC DR PROSITE; PS00348; P53; 1.
CC DR PFAM; PF00870; P53; 1.
CC KW Anti-oncogene; DNA-binding; Transcription regulation; Activator;
CC Nuclear protein; Phosphorylation; Apoptosis.
CC FT DOMAIN 1 74 ASP/GLU-RICH (ACIDIC).
CC FT DOMAIN 75 150 HYDROPHOBIC.
CC FT DOMAIN 316 390 HIGHLY BASIC AND MAY BE INVOLVED IN
CC INTERACTION WITH DNA.
CC FT DOMAIN 311 323 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
CC FT MOD_RES 392 392 PHOSPHORYLATION (BY SIMILARITY).
CC FT VARIANT 133 133 L -> Q (IN CELL LINE V79-4).
CC FT VARIANT 135 135 C -> W (IN CELL LINE V79-4).
CC FT CONFLICT 103 103 Y -> F (IN REF. 2).
CC SQ SEQUENCE 393 AA; 43378 MW; 402EB149 CRC32;

Query Match 89.7%; Score 70; DB 1; Length 393;
Best Local Similarity 90.0%; Pred. No. 5.93e-04;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 321 KTLDDGEYFTL 330
QY 1 KPLDGEYFTL 10

Search completed: Sat Apr 15 01:06:34 2000
Job time : 40 secs.
```

```
CC BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2
CC EXPRESSION (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY
CC OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED
CC IN MANY TYPES OF CANCER.
CC -!- SIMILARITY: BELONGS TO THE P53 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U26741; ABA41265.1; -
CC HSSP: P04637; ITR.
CC DR PROSITE: PS00348; P53; 1.
CC DR PFAM: PF00870; P53; 1.
CC KW Anti-oncogene; DNA-binding; Transcription regulation; Activator;
CC Nuclear protein; Phosphorylation; Apoptosis.
CC FT NON_TER 1 199 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
CC FT DOMAIN 187 207
CC FT NON_TER 207 207
CC SQ SEQUENCE 207 AA; 23428 MW; 0FBAB9C1 CRC32;
Query Match 91.0%; Score 71; DB 1; Length 207;
Best Local Similarity 100.0%; Pred. No. 3.39e-04;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 198 PLDGEYFTL 206
QY 2 PLDGEYFTL 10
|||||||
RESULT 14
ID P53_HORSE STANDARD; PRT; 280 AA.
AC P79892; Q29481;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).
GN TP53 OR P53.
OS Equus caballus (Horse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Perissodactyla; Equidae; Equus.
RN [1]
RP SEQUENCE OF 1-263 FROM N.A.
RC TISSUE=SPLEEN;
RX MEDLINE; 97070350.
RA PAZZI K.A., KRAEGL S.A., GRIFFEY S.M., THEON A.P., MADEWELL B.R.;
RT "Analysis of the equine tumor suppressor gene p53 in the normal horse
RT and in eight cutaneous squamous cell carcinomas.";
RL Cancer Lett. 107:125-130(1996).
RN [2]
RP SEQUENCE OF 76-280 FROM N.A.
RX MEDLINE; 96293865.
RA NASIR L., REID S.W.;
RT "Nucleotide sequence of exons 5 to 9 of the p53 tumour-suppressor
RT gene of the horse (Equus caballus).";
RL DNA Seq. 6:185-187(1996).
CC -!- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES
CC GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL
CC CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN
CC TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION, IT IS A
CC TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION
CC BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF
CC THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.
CC APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF
CC BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2
CC EXPRESSION (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
```

```
CC -!- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY
CC OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED
CC IN MANY TYPES OF CANCER.
CC -!- SIMILARITY: BELONGS TO THE P53 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: S83123; AAB46899.1; -
CC EMBL: U37120; AAB18936.1; -
CC HSSP: P04637; ISAH.
CC DR PROSITE: PS00348; P53; 1.
CC DR PFAM: PF00870; P53; 1.
CC KW Anti-oncogene; DNA-binding; Transcription regulation; Activator;
CC Nuclear protein; Phosphorylation; Apoptosis.
CC FT NON_TER 1 274 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
CC FT DOMAIN 262 274
CC FT CONFLICT 79 79 T -> A (IN REF. 2).
CC FT CONFLICT 83 83 L -> M (IN REF. 2).
CC FT CONFLICT 111 111 A -> V (IN REF. 2).
CC FT CONFLICT 138 138 G -> A (IN REF. 2).
CC FT NON_TER 280 280
CC SQ SEQUENCE 280 AA; 30985 MW; B494F872 CRC32;
Query Match 89.7%; Score 70; DB 1; Length 280;
Best Local Similarity 100.0%; Pred. No. 5.93e-04;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 272 KPLDGEYFT 280
QY 1 KPLDGEYFT 9
|||||||
RESULT 15
ID P53_CRIGR STANDARD; PRT; 393 AA.
AC O09185; Q64397; P97258; P97788;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE CELLULAR TUMOR ANTIGEN P53.
GN TP53 OR P53.
OS Cricetus griseus (Chinese hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae; Cricetulus.
RN [1]
RP SEQUENCE FROM N.A.
RA CHUNG W., MI L.J., BOORSTEIN R.J.;
RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RX MEDLINE; 97183659.
RA LEE H., LARNER J.M., HAMLIN J.L.;
RT "Cloning and characterization of Chinese hamster p53 cDNA.";
RL Gene 184:177-183(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=EMBRYONIC FIBROBLAST;
RA SHIMIZU T., NIKAI DO O., SUZUKI F.;
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES
CC GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL
CC CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN
CC TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION, IT IS A
CC TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION
CC BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF
CC THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.
CC APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF
CC BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2
CC EXPRESSION (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
```



OC Eutheria; Primates; Catarrhini; Cercopithecidae; Cercopithecinae;  
OC Chlorocebus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=LIVER;  
RX MEDLINE; 90045967.  
RA RIGAUDY P., ECKHARDT W.;  
RT "Nucleotide sequence of a cDNA encoding the monkey cellular  
phosphoprotein p53.";  
RL Nucleic Acids Res. 17:8375-8375(1989).  
CC -1- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES  
GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL  
CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN  
TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION, IT IS A  
TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION  
BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF  
THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.  
CC APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF  
BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2  
EXPRESSION.  
CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
CC -1- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY  
OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED  
IN MANY TYPES OF CANCER.  
CC -1- SIMILARITY: BELONGS TO THE P53 FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>;  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; X16384; CAA34420.1; -.  
DR PIR; S06594; S06594.  
DR HSP; P04637; 1SAH.  
DR PROSITE; P500348; P53; 1.  
DR PFAM; PF00870; P53; 1.  
KW Anti-oncogene; DNA-binding; Transcription regulation; Activator;  
KW Nuclear protein; Phosphorylation; Apoptosis.  
FT DOMAIN 1 68  
FT DOMAIN 81 150  
FT DOMAIN 319 393  
FT HIGHLY BASIC AND MAY BE INVOLVED IN  
FT INTERACTION WITH DNA.  
FT DOMAIN 311 323  
FT NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
FT MOD\_RES 392 392  
FT PHOSPHORYLATION (BY SIMILARITY).  
SQ SEQUENCE 393 AA; 43696 MW; BBE7DC62 CRC32;  
Query Match 100.0%; Score 78; DB 1; Length 393;  
Best Local Similarity 100.0%; Pred. No. 6.16e-06;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Db 321 KPLDGEYFTL 330  
QY |||||  
1 KPLDGEYFTL 10  
RESULT 12  
ID P53\_RABIT STANDARD; PRT; 391 AA.  
AC Q95330;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE CELLULAR TUMOR ANTIGEN P53.  
GN TP53.  
OS Oryctolagus cuniculus (Rabbit).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Lagomorpha; Leporidae; Oryctolagus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NEW ZEALAND;  
RX MEDLINE; 97208869.

RA LE GOAS F., MAY P., RONCO P., CARON DE FROMENTEL C.;  
RT "cDNA cloning and immunological characterization of rabbit p53.";  
RL Gene 185:169-173(1997).  
CC -1- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES  
GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL  
CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN  
TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION, IT IS A  
TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION  
BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF  
THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.  
CC APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF  
BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2  
EXPRESSION (BY SIMILARITY).  
CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
CC -1- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY  
OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED  
IN MANY TYPES OF CANCER.  
CC -1- SIMILARITY: BELONGS TO THE P53 FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>;  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; X90592; CAA62216.1; -.  
DR HSP; P04637; 1YCR.  
DR PROSITE; P500348; P53; 1.  
DR PFAM; PF00870; P53; 1.  
KW Anti-oncogene; DNA-binding; Transcription regulation; Activator;  
KW Nuclear protein; Phosphorylation; Apoptosis.  
FT DOMAIN 1 70  
FT DOMAIN 308 321  
FT NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
FT MOD\_RES 390 390  
FT PHOSPHORYLATION (BY SIMILARITY).  
SQ SEQUENCE 391 AA; 43435 MW; 30A36172 CRC32;  
Query Match 93.6%; Score 73; DB 1; Length 391;  
Best Local Similarity 90.0%; Pred. No. 1.10e-04;  
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Db 319 KPLDGEYFTL 328  
QY |||||  
1 KPLDGEYFTL 10  
RESULT 13  
ID P53\_EQUUS STANDARD; PRT; 207 AA.  
AC Q29480;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).  
GN TP53.  
OS Equus asinus (Donkey).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Perissodactyla; Equidae; Equus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 96342529.  
RA NASIR L., REID S.W.;  
RT "Nucleotide sequence of exons 5 to 9 of the p53 tumour-suppressor  
gene of the donkey (Equus asinus).";  
RL DNA Seq. 6:61-63(1995).  
CC -1- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES  
GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL  
CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN  
TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION, IT IS A  
TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION  
BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF  
THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.  
CC APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF

RN [19] VARIANT ARG-72.  
 RX MEDLINE; 91153807.  
 RA OLSCHWANG S., LAURENT-PUIG P., VASSAL A., SALMON R.-J., THOMAS G.;  
 RT "Characterization of a frequent polymorphism in the coding sequence  
 of the p53 gene in colonic cancer patients and a control  
 population.";  
 RL Hum. Genet. 86:369-370(1991).  
 RN [20]  
 RP VARIANT LFS THR-133.  
 RX MEDLINE; 92034774.  
 RA LAW J.C., STRONG L.C., CHIDAMBARAM A., FERRELL R.E.;  
 RT "A germ line mutation in exon 5 of the p53 gene in an extended cancer  
 family.";  
 RL Cancer Res. 51:6385-6387(1991).  
 RN [21]  
 RP VARIANTS LFS CYS-245; TRP-248; PRO-252 AND LYS-258.  
 RX MEDLINE; 91057857.  
 RA MALKIN D., LI F.P., STRONG L.C., FRAUMENI J.F. JR., NELSON C.E.,  
 RA KIM D.H., KASSEL J., GRYKA M.A., BISCHOFF F.Z., TAINSKY M.A.,  
 RA FRIEND S.H.;  
 RT "Germ line p53 mutations in a familial syndrome of breast cancer,  
 sarcomas, and other neoplasms.";  
 RL Science 250:1233-1238(1990).  
 RN [22]  
 RP VARIANT LFS ASP-245.  
 RX MEDLINE; 91080929.  
 RA SRIVASTAVA S., ZOU Z., PIROLLO K., BLATTNER W., CHANG E.H.;  
 RT "Germ-line transmission of a mutated p53 gene in a cancer-prone  
 family with Li-Fraumeni syndrome.";  
 RL Nature 348:747-749(1990).  
 RN [23]  
 RP VARIANT LFS LEU-272.  
 RX MEDLINE; 92147883.  
 RA FELIX C.A., NAU M.M., TAKAHASHI T., MITSUDOMI T., CHIBA I.,  
 RA POPLACK D.G., REAMAN G.H., COLE D.E., LETTERIO J.J., WHANG-PENG J.,  
 RA NUTSEN T., MINNA J.D.;  
 RT "Hereditary and acquired p53 gene mutations in childhood acute  
 lymphoblastic leukemia.";  
 RL J. Clin. Invest. 89:640-647(1992).  
 RN [24]  
 RP VARIANTS LFS HIS-273 AND VAL-325.  
 RX MEDLINE; 92228023.  
 RA MALKIN D., JOLLY K.W., BARBIER N., LOOK A.T., FRIEND S.H.,  
 RA GEBHARDT M.C., ANDERSEN T.I., BORRESEN A.-L., LI F.P., GARBER J.,  
 RA STRONG L.C.;  
 RT "Germline mutations of the p53 tumor-suppressor gene in children and  
 young adults with second malignant neoplasms.";  
 RL New Engl. J. Med. 326:1309-1315(1992).  
 RN [25]  
 RP VARIANTS BREAST TUMORS GLN-132; SER-249; LYS-280 AND LYS-285.  
 RX MEDLINE; 90295284.  
 RA BARTKE J., IGGO R., GANNON J., LANE D.P.;  
 RT "Genetic and immunochemical analysis of mutant p53 in human breast  
 cancer cell lines.";  
 RL Oncogene 5:893-899(1990).  
 RN [26]  
 RP VARIANTS COLON TUMORS PHE-241 AND HIS-273.  
 RX MEDLINE; 91017544.  
 RA RODRIGUES N.R., ROWAN A., SMITH M.E.F., KERR I.B., BODMER W.F.,  
 RA GANNON J.V., LANE D.P.;

....  
 Note: remainder of annotations omitted.

Query Match 100.0%; Score 78; DB 1; Length 393;  
 Best Local Similarity 100.0%; Pred. No. 6.16e-06;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 321 KPLDGEYFTL 330  
 |||||  
 QY 1 KPLDGEYFTL 10

RESULT 10  
 ID P53\_MACFA STANDARD; PRT; 393 AA.  
 AC P56423;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE CELLULAR TUMOR ANTIGEN P53.  
 GN TP53 OR P53.  
 OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Primates; Catarrhini; Cercopithecoidea; Cercopithecinae;  
 OC Macaca.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA KHAN M.A., HANSEN C., WELSH J.A., BENNETT W.P.;  
 RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.  
 CC -!- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES  
 CC GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL  
 CC CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN  
 CC TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION, IT IS A  
 CC TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION  
 CC BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF  
 CC THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.  
 CC APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF  
 CC BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2  
 CC EXPRESSION.  
 CC -!- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -!- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY  
 CC OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED  
 CC IN MANY TYPES OF CANCER.  
 CC -!- SIMILARITY: BELONGS TO THE P53 FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; U48957; AAB91535.1; -  
 DR HSSP; P04637; ISAH.  
 DR PROSITE; PS00348; P53; 1.  
 DR PFAM; PF00870; P53; 1.  
 KW Anti-oncogene; DNA-binding; Transcription regulation; Activator;  
 KW Nuclear protein; Phosphorylation; Apoptosis  
 FT DOMAIN 1 80 ASP/GLU-RICH (ACIDIC).  
 FT DOMAIN 81 150 HYDROPHOBIC.  
 FT DOMAIN 319 393 HIGHLY BASIC AND MAY BE INVOLVED IN  
 FT INTERACTION WITH DNA.  
 FT DOMAIN 311 323 NUCLEAR LOCALIZATION SIGNAL.  
 FT MOD\_RES 392 392 PHOSPHORYLATION (BY SIMILARITY).  
 SQ SEQUENCE 393 AA; 43678 MW; 2499AC47 CRC32;  
 Query Match 100.0%; Score 78; DB 1; Length 393;  
 Best Local Similarity 100.0%; Pred. No. 6.16e-06;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 321 KPLDGEYFTL 330  
 |||||  
 QY 1 KPLDGEYFTL 10  
 RESULT 11  
 ID P53\_CERAE STANDARD; PRT; 393 AA.  
 AC P13481;  
 DT 01-JAN-1990 (Rel. 13, Created)  
 DT 01-JAN-1990 (Rel. 13, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE CELLULAR TUMOR ANTIGEN P53.  
 GN TP53.  
 OS Cercopithecus aethiops (Green monkey) (Grivet).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

DB 321 KPLDGEYFTL 330  
QY 1 KPLDGEYFTL 10  
RESULT 9  
ID P53\_HUMAN STANDARD; PRT; 393 AA.  
AC P04637;  
DT 13-AUG-1987 (Rel. 05, Created)  
DT 01-MAR-1989 (Rel. 10, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE CELLULAR TUMOR ANTIGEN P53 (PHOSPHOPROTEIN P53).  
GN TP53.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 85230577.  
RA ZAKUT-HOURI R., BIENZ-TADMOR B., GIVOL D., OREN M.;  
RT "Human p53 cellular tumor antigen: cDNA sequence and expression in  
RT COS cells.";  
RL EMBO J. 4:1251-1255(1985).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 87064416.  
RA LAMB P., CRAWFORD L.;  
RT "Characterization of the human p53 gene.";  
RL Mol. Cell. Biol. 6:1379-1385(1986).  
RN [3]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 85267676.  
RA HARLOW E., WILLIAMSON N.M., RALSTON R., HELFMAN D.M., ADAMS T.E.;  
RT "Molecular cloning and in vitro expression of a cDNA clone for human  
RT cellular tumor antigen p53.";  
RL Mol. Cell. Biol. 5:1601-1610(1985).  
RN [4]  
RP TRANSFORMED HYBRIDOMA SV-80 CELL LINE, SEQUENCE FROM N.A.  
RX MEDLINE; 87089826.  
RA HARRIS N., BRILL E., SHOHAT O., PROKOCIMER M., WOLF D., ARAI N.,  
RA ROTTIER V.;  
RT "Molecular basis for heterogeneity of the human p53 protein.";  
RL Mol. Cell. Biol. 6:4650-4656(1986).  
RN [5]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 89108008.  
RA BUCHANAN V.L., CHURAKOV P.M., NINKINA N.N., SAMARINA O.P.,  
RA GEORGIEV G.P.;  
RT "A variation in the structure of the protein-coding region of the  
RT human p53 gene.";  
RL Gene 70:245-252(1988).  
RN [6]  
RP SEQUENCE OF 101-393 FROM N.A.  
RX MEDLINE; 85126934.  
RA MATLASHESKI G., LAMB P., PIM D., PEACOCK J., CRAWFORD L.,  
RA BENCHIMOL S.;  
RT "Isolation and characterization of a human p53 cDNA clone: expression  
RT of the human p53 gene.";  
RL EMBO J. 3:3257-3262(1984).  
RN [7]  
RP NUCLEAR LOCALIZATION SIGNAL.  
RX MEDLINE; 90191730.  
RA ADDISON C., JENKINS J.R., STURZBECHER H.-W.;  
RT "The p53 nuclear localisation signal is structurally linked to a  
RT p34cdc2 kinase motif.";  
RL Oncogene 5:423-426(1990).  
RN [8]  
RP PHOSPHORYLATION BY P60/CDC2 AND CYCLIN B/CDC2.  
RX MEDLINE; 90280456.  
RA BITSCHOFF J.R., FRIEDMAN P.N., MARSHAK D.R., PRIVES C., BEACH D.;  
RT "Human p53 is phosphorylated by p60-cdc2 and cyclin B-cdc2.";  
RT Proc. Natl. Acad. Sci. U.S.A. 87:4766-4770(1990).  
RN [9]  
RP DEPHOSPHORYLATION BY PP2A.  
RX MEDLINE; 91172186.  
RA SCHEIDTMANN K.H., MUMBY M.C., RUNDELL K., WALTER G.;  
RT "Dephosphorylation of simian virus 40 large-T antigen and p53 protein  
RT by protein phosphatase 2A: inhibition by small-t antigen.";  
RL Mol. Cell. Biol. 11:1996-2003(1991).  
RN [10]  
RP STRUCTURE BY NMR OF 319-360.  
RX MEDLINE; 94294808.  
RA CLORE G.M., OMICHINSKI J.G., SAKAGUCHI K., ZAMBRANO N., SAKAMOTO H.,  
RA APPELLA E., GRONENBORN A.M.;  
RT "High-resolution structure of the oligomerization domain of p53 by  
RT multidimensional NMR.";  
RL Science 265:386-391(1994).  
RN [11]  
RP STRUCTURE BY NMR OF 325-355.  
RX MEDLINE; 95292092.  
RA LEE W., HARVEY T.S., YIN Y., YAU P., LITCHFIELD D., ARROWSMITH C.H.;  
RT "Solution structure of the tetrameric minimum transforming domain of  
RT p53.";  
RL Nat. Struct. Biol. 1:877-890(1994).  
RN [12]  
RP STRUCTURE BY NMR OF 326-354.  
RX MEDLINE; 98026899.  
RA MCCOY M., STAVRIDI E.S., WATERMAN J.L., WIECZOREK A.M., OPELLA S.J.,  
RA HALAZONETIS T.D.;  
RT "Hydrophobic side-chain size is a determinant of the  
RT three-dimensional structure of the p53 oligomerization domain.";  
RL EMBO J. 16:6230-6236(1997).  
RN [13]  
RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 94-289.  
RX MEDLINE; 94294806.  
RA CHO Y., GORINA S., JEFFREY P.D., PAVLETICH N.P.;  
RT "Crystal structure of a p53 tumor suppressor-DNA complex:  
RT understanding tumorigenic mutations.";  
RL Science 265:346-355(1994).  
RN [14]  
RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS) OF 13-29 IN COMPLEX WITH MDM2.  
RX MEDLINE; 97081050.  
RA KUSSIE P.H., GORINA S., MARECHAL V., ELENBAAS B., MOREAU J.,  
RA LEVINE A.J., PAVLETICH N.P.;  
RT "Structure of the MDM2 oncoprotein bound to the p53 tumor suppressor  
RT transactivation domain.";  
RL Science 274:948-953(1996).  
RN [15]  
RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 97-287 IN COMPLEX WITH 53BP2.  
RX MEDLINE; 97035414.  
RA GORINA S., PAVLETICH N.P.;  
RT "Structure of the p53 tumor suppressor bound to the ankyrin and SH3  
RT domains of 53BP2.";  
RL Science 274:1001-1005(1996).  
RN [16]  
RP REVIEW.  
RX MEDLINE; 94090335.  
RA HARRIS C.C.;  
RT "p53: at the crossroads of molecular carcinogenesis and risk  
RT assessment.";  
RL Science 262:1980-1981(1993).  
RN [17]  
RP REVIEW ON VARIANTS.  
RX MEDLINE; 91289156.  
RA HOOLSTEIN M., SIDRANSKY D., VOGELSTEIN B., HARRIS C.C.;  
RT "p53 mutations in human cancers.";  
RL Science 253:49-53(1991).  
RN [18]  
RP REVIEW ON VARIANTS.  
RX MEDLINE; 96271983.  
RA DE VRIES E.M.G., RICHE D.O., DE VRIES T.N., HARTMANN A., BLASZYK H.,  
RA LIAO D., SOUSSI T., KOVACH J.S., SOMMER S.S.;  
RT "Database of mutations in the p53 and APC tumor suppressor genes  
RT designed to facilitate molecular epidemiological analyses.";  
RL Hum. Mutat. 7:202-213(1996).

```

RESULT 7
ID P53_RAT STANDARD; PRT; 391 AA.
AC P10161: 009168;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE CELLULAR TUMOR ANTIGEN P53.
GN TP53 OR P53.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 89083585.
RA SOUSSI T.;
RT "Nucleotide sequence of a cDNA encoding the rat p53 nuclear
RT oncoprotein."
RL Nucleic Acids Res. 16:11384-11384(1988).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE: 93181268.
RA HULLA J.E., SCHNEIDER R.P.;
RT "Structure of the rat p53 tumor suppressor gene."
RL Nucleic Acids Res. 21:713-717(1993).
RN [3]
RP SEQUENCE FROM N.A.
RX STRAIN-SPRAGUE-DAWLEY;
RA MATHUPALA S.P.;
RT Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES
CC GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL
CC CIRCUMSTANCES OR CELL TYPE. BUT BOTH ACTIVITIES ARE INVOLVED IN
CC TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION. IT IS A
CC TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION
CC BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF
CC THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.
CC APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF
CC BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2
CC EXPRESSION.
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY
CC OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED
CC IN MANY TYPES OF CANCER.
CC -1- SIMILARITY: BELONGS TO THE P53 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: X13058; CAA31457.1;
CC EMBL: L07910; AAA41788.1;
CC EMBL: L07910; AAA41788.1; JOINED.
CC EMBL: L07904; AAA41788.1; JOINED.
CC EMBL: L07905; AAA41788.1; JOINED.
CC EMBL: L07906; AAA41788.1; JOINED.
CC EMBL: L07907; AAA41788.1; JOINED.
CC EMBL: L07908; AAA41788.1; JOINED.
CC EMBL: L07909; AAA41788.1; JOINED.
CC EMBL: U90328; AAB80959.1;
CC PIR: S02192; S02192.
CC HSSP: P04637; 1PBT.
CC PROSITE: PS00348; P53; 1.
CC PFAM: PF00870; P53; 1.
CC Anti-oncogene; DNA-binding; Transcription regulation; Activator;
CC Nuclear protein; Phosphorylation; Apoptosis.
CC DOMAIN 1 76 ASP/GLU-RICH (ACIDIC).
CC DOMAIN 277 151 HYDROPHOBIC.
CC DOMAIN 277 391 HIGHLY BASIC AND MAY BE INVOLVED IN
CC INTERACTION WITH DNA.
CC
CC Query Match 100.0%; Score 78; DB 1; Length 391;
CC Best Local Similarity 100.0%; Pred. No. 6.16e-06;
CC Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

FT DOMAIN 309 321 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT MOD_RES 390 390 PHOSPHORYLATION (BY SIMILARITY).
FT VARIANT 103 103 G -> S.
FT VARIANT 256 256 E -> G.
FT CONFLICT 174 174 C -> W (IN REF. 2).
SQ SEQUENCE 391 AA; 43451 MW; E0114C18 CRC32;

Query Match 100.0%; Score 78; DB 1; Length 391;
Best Local Similarity 100.0%; Pred. No. 6.16e-06;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 319 KPLDGEYFTL 328
QY 1 KPLDGEYFTL 10

RESULT 8
ID P53_MACMU STANDARD; PRT; 393 AA.
AC P56424;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE CELLULAR TUMOR ANTIGEN P53.
GN TP53 OR P53.
OS Macaca mulatta (Rhesus macaque).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Cercopitheciinae; Cercopithecinae;
OC Macaca.
RN [1]
RP SEQUENCE FROM N.A.
RA KHAN M.A., HANSEN C., WELSH J.A., BENNETT W.P.;
RT Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES
CC GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL
CC CIRCUMSTANCES OR CELL TYPE. BUT BOTH ACTIVITIES ARE INVOLVED IN
CC TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION. IT IS A
CC TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION
CC BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF
CC THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.
CC APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF
CC BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2
CC EXPRESSION.
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY
CC OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED
CC IN MANY TYPES OF CANCER.
CC -1- SIMILARITY: BELONGS TO THE P53 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U48956; AAB91534.1;
CC HSSP: P04637; 1SAH.
CC PROSITE: PS00348; P53; 1.
CC PFAM: PF00870; P53; 1.
CC Anti-oncogene; DNA-binding; Transcription regulation; Activator;
CC Nuclear protein; Phosphorylation; Apoptosis.
CC DOMAIN 1 80 ASP/GLU-RICH (ACIDIC).
CC DOMAIN 81 150 HYDROPHOBIC.
CC DOMAIN 319 393 HIGHLY BASIC AND MAY BE INVOLVED IN
CC INTERACTION WITH DNA.
CC
CC DOMAIN 311 323 NUCLEAR LOCALIZATION SIGNAL.
FT MOD_RES 392 392 PHOSPHORYLATION (BY SIMILARITY).
SQ SEQUENCE 393 AA; 43655 MW; 11A9B7F8 CRC32;

Query Match 100.0%; Score 78; DB 1; Length 393;
Best Local Similarity 100.0%; Pred. No. 6.16e-06;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```



Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 310 KPLDGEYFTL 319  
|||||  
QY 1 KPLDGEYFTL 10

RESULT 4  
ID P53\_FELCA STANDARD; PRT; 386 AA.  
AC P41585;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE CELLULAR TUMOR ANTIGEN P53.  
GN TP53.  
OS Felis silvestris catus (Cat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Carnivora; Fissipedia; Felidae; Felis.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-LYMPH NODE;  
RX MEDLINE; 94333960.  
RA OKUDA M., UEDA A., SAKAI T., OHASHI T., MOMOI Y., YOUN H.Y.,  
WATARI T., GOITSUKA R., TSUJIMOTO H., HASEGAWA A.;  
RT "Cloning of feline p53 tumor-suppressor gene and its aberration in  
hematopoietic tumors."  
RL Int. J. Cancer 58:602-607(1994).  
RN [2]  
RP SEQUENCE OF 34-354 FROM N.A.  
RX MEDLINE; 94114699.  
RA OKUDA M., UEDA A., MATSUMOTO Y., MOMOI Y., WATARI T., GOITSUKA R.,  
O'BRIEN S.J., TSUJIMOTO H., HASEGAWA A.;  
RT "Molecular cloning and chromosomal mapping of feline p53 tumor  
suppressor gene."  
RL J. Vet. Med. Sci. 55:801-805(1993).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC SPECIES-BOVINE; STRAIN-HOLSTEIN; TISSUE-THYMUS;  
RX MEDLINE; 96401400.  
RA KOMORI H., ISHIGURO N., HORIUCHI M., SHINAGAWA M., AIDA Y.;  
RT "Predominant p53 mutations in enzootic bovine leukemic cell lines."  
RL Vet. Immunol. Immunopathol. 52:53-63(1996).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC SPECIES-B. INDICUS; STRAIN-BORAN; TISSUE-BLOOD;  
RX BISHOP R.R.P., GOBRIGHT E.E.I.;  
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES  
GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL  
CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN  
TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION, IT IS A  
TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION  
BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF  
THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.  
APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF  
BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2  
EXPRESSION.  
CC -!- SUBCELLULAR LOCATION: NUCLEAR.  
CC -!- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY  
OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED  
IN MANY TYPES OF CANCER.  
CC -!- SIMILARITY: BELONGS TO THE P53 FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; D26608; BAA05653.1; -;  
CC EMBL; D16460; BAA03927.1; -;  
CC HSSP; P04637; 1SAH.  
CC PROSITE; PS00348; P53; 1.  
CC PFAM; PF00870; P53; 1.  
CC Anti-oncogene; DNA-binding; Transcription regulation; Activator;  
KW Nuclear protein; Phosphorylation; Apoptosis.  
FT DOMAIN 1 59 ASP/GLU-RICH (ACIDIC).  
FT MOD\_RES 304 316 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
FT FT 385 PHOSPHORYLATION (BY SIMILARITY).  
FT FT 285 K -> R (IN REF. 2).  
SQ SEQUENCE 386 AA; 42692 MW; D6C7132A CRC32;  
Query Match 100.0%; Score 78; DB 1; Length 386;  
Best Local Similarity 100.0%; Pred. NO. 6.16e-06;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 314 KPLDGEYFTL 323  
|||||  
QY 1 KPLDGEYFTL 10

RESULT 5  
ID P53\_BOVIN STANDARD; PRT; 386 AA.  
AC Q29628;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 15-JUL-1998 (Rel. 36, Last annotation update)  
DE CELLULAR TUMOR ANTIGEN P53.  
GN TP53.  
OS Bos taurus (Bovine), and Bos indicus (Zebu).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
OC Bovinae; Bos.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC SPECIES-BOVINE; TISSUE-LIVER;  
RX MEDLINE; 95352829.  
RA DEQUIEDT F., KETTMANN R., BURNY A., WILLEMS L.;  
RT "Nucleotide sequence of the bovine P53 tumor-suppressor cDNA."  
RL DNA Seq. 5:261-264(1995).  
RN [2]  
RP SEQUENCE OF 13-386 FROM N.A.  
RC SPECIES-BOVINE; STRAIN-HOLSTEIN; TISSUE-THYMUS;  
RX MEDLINE; 96401400.  
RA KOMORI H., ISHIGURO N., HORIUCHI M., SHINAGAWA M., AIDA Y.;  
RT "Predominant p53 mutations in enzootic bovine leukemic cell lines."  
RL Vet. Immunol. Immunopathol. 52:53-63(1996).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC SPECIES-B. INDICUS; STRAIN-BORAN; TISSUE-BLOOD;  
RX BISHOP R.R.P., GOBRIGHT E.E.I.;  
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES  
GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL  
CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN  
TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION, IT IS A  
TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION  
BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF  
THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.  
APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF  
BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2  
EXPRESSION.  
CC -!- SUBCELLULAR LOCATION: NUCLEAR.  
CC -!- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY  
OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED  
IN MANY TYPES OF CANCER.  
CC -!- SIMILARITY: BELONGS TO THE P53 FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; X81704; CAA57348.1; -;  
CC EMBL; D49825; BAA08629.1; -;  
CC EMBL; U74486; AAB51214.1; -;  
CC HSSP; P04637; 1YCR.  
CC PROSITE; PS00348; P53; 1.  
CC PFAM; PF00870; P53; 1.  
CC Anti-oncogene; DNA-binding; Transcription regulation; Activator;  
KW Nuclear protein; Phosphorylation; Apoptosis.  
FT DOMAIN 1 59 ASP/GLU-RICH (ACIDIC).  
FT FT 316 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
FT FT 385 PHOSPHORYLATION (BY SIMILARITY).  
FT MOD\_RES 304 316 ASP/GLU-RICH (ACIDIC).  
FT FT 385 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
FT FT 385 PHOSPHORYLATION (BY SIMILARITY).  
FT MOD\_RES 385

```

KW Nuclear protein; Phosphorylation; Apoptosis.
FT DOMAIN 1 59 ASP/GLU-RICH (ACIDIC).
FT DOMAIN 68 137 HYDROPHOBIC.
FT DOMAIN 307 381 HIGHLY BASIC AND MAY BE INVOLVED IN
FT INTERACTION WITH DNA (BY SIMILARITY).
FT FT INTERACT WITH DNA (BY SIMILARITY).
FT FT NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT FT PHOSPHORYLATION (BY SIMILARITY).
SQ SEQUENCE 381 AA; 42486 MW; 70210B63 CRC32;

Query Match 100.0%; Score 78; DB 1; Length 381;
Best Local Similarity 100.0%; Pred. No. 6.16e-06;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 309 KPLDGEYFTL 318
      |||||
QY 1 KPLDGEYFTL 10

RESULT 3
ID P53_SHEEP STANDARD; PRT; 382 AA.
AC P51664;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE CELLULAR TUMOR ANTIGEN P53.
DE TP53.
GN Ovis aries (Sheep).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae;
OC Caprinae; Ovis.
OC [1]
RN SEQUENCE FROM N.A.
RP.
RC TISSUE-BLOOD;
RX MEDLINE; 95352828.
RA DEQUEDT F., KETTMANN R., BURNY A., WILLEMS L.;
RT "Nucleotide sequence of the ovine P53 tumor-suppressor cDNA and its
RT genomic organization.";
RL DNA Seq. 5:1255-259(1995).
CC -1- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES
CC GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL
CC CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN
CC TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION, IT IS A
CC TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION
CC BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF
CC THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.
CC APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF
CC BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2
CC EXPRESSION.
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- SIMILARITY: BELONGS TO THE P53 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X81705; CAA57349.1; -
CC DR HSP; P04637; 1PET.
CC DR PROSITE; PS00348; P53; 1.
CC DR PFAM; PF00870; P53; 1.
CC KW Anti-oncogene; DNA-binding; Transcription regulation; Activator;
CC FT Nuclear protein; Phosphorylation; Apoptosis.
CC FT DOMAIN 1 66 ASP/GLU-RICH (ACIDIC).
CC FT DOMAIN 300 312 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
CC FT MOD_RES 381 381 PHOSPHORYLATION (BY SIMILARITY).
CC SQ SEQUENCE 382 AA; 42809 MW; 0CB99A00 CRC32;

Query Match 100.0%; Score 78; DB 1; Length 382;
Best Local Similarity 100.0%; Pred. No. 6.16e-06;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```



\*\*\*\*\*

W P E L L (TM)

\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Sat Apr 15 01:05:54 2000; MasPar time 3.12 Seconds  
Tabular output not generated. 95.591 Million cell updates/sec

Title: >US-08-452-843-20  
Description: (1-10) from US08452843.pep  
Perfect Score: 78  
Sequence: 1 KPLDGEYFTL 10  
  
Scoring table: PAM 150  
Gap 15  
  
Searched: 82229 seqs, 29864866 residues  
  
Post-processing: Minimum Match 0%  
Listing first 45 summaries  
  
Database: swiss-prot38  
1:swissprot  
  
Statistics: Mean 25.637; Variance 30.391; scale 0.844

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	78	100.0	314	1 P53_SPEBE	CELLULAR TUMOR ANTIGEN	6.16e-06
2	78	100.0	381	1 P53_CANFA	CELLULAR TUMOR ANTIGEN	6.16e-06
3	78	100.0	382	1 P53_SHEEP	CELLULAR TUMOR ANTIGEN	6.16e-06
4	78	100.0	386	1 P53_FELCA	CELLULAR TUMOR ANTIGEN	6.16e-06
5	78	100.0	386	1 P53_BOVIN	CELLULAR TUMOR ANTIGEN	6.16e-06
6	78	100.0	390	1 P53_MOUSE	CELLULAR TUMOR ANTIGEN	6.16e-06
7	78	100.0	391	1 P53_RAT	CELLULAR TUMOR ANTIGEN	6.16e-06
8	78	100.0	393	1 P53_MACMU	CELLULAR TUMOR ANTIGEN	6.16e-06
9	78	100.0	393	1 P53_HUMAN	CELLULAR TUMOR ANTIGEN	6.16e-06
10	78	100.0	393	1 P53_MACFA	CELLULAR TUMOR ANTIGEN	6.16e-06
11	78	100.0	393	1 P53_CERAE	CELLULAR TUMOR ANTIGEN	6.16e-06
12	73	93.6	391	1 P53_RABIT	CELLULAR TUMOR ANTIGEN	1.10e-04
13	71	91.0	207	1 P53_EQUAS	CELLULAR TUMOR ANTIGEN	3.39e-04
14	70	89.7	280	1 P53_HORSE	CELLULAR TUMOR ANTIGEN	5.93e-04
15	70	89.7	393	1 P53_CRIGR	CELLULAR TUMOR ANTIGEN	5.93e-04
16	70	89.7	396	1 P53_MESAU	CELLULAR TUMOR ANTIGEN	5.93e-04
17	56	71.8	499	1 DHAS_CHICK	RETINALDEHYDE-SPECIFIC	9.12e-01
18	56	71.8	499	1 DHAS_HUMAN	RETINALDEHYDE-SPECIFIC	9.12e-01
19	56	71.8	499	1 DHAS_RAT	RETINALDEHYDE-SPECIFIC	9.12e-01
20	56	71.8	499	1 DHAS_MOUSE	RETINALDEHYDE-SPECIFIC	9.12e-01
21	54	69.2	349	1 ASPX_VULVU	SPERM ACROSOMAL PROTEIN	2.39e+00
22	54	69.2	382	1 MATB_NEUCR	MATING TYPE PROTEIN A	2.39e+00
23	54	69.2	640	1 Y551_SYNY3	HYPOTHETICAL 70.4 KD P	2.39e+00

24	54	69.2	694	1	LCF4_YEAST	LONG-CHAIN-FATTY-ACID-LCF3_YEAST	2.39e+00
25	54	69.2	694	1	LCF3_YEAST	LONG-CHAIN-FATTY-ACID-AAC-RICH MRNA CLONE AA	2.39e+00
26	53	67.9	183	1	ADAM_BOVIN	ALDEHYDE DEHYDROGENASE	3.83e+00
27	53	67.9	520	1	ET1B_XENLA	C-ETS-1B PROTEIN (XEL-6.10e+00	6.10e+00
28	52	66.7	268	1	ET1B_XENLA	C-ETS-1A PROTEIN	6.10e+00
29	52	66.7	438	1	FLT3_MOUSE	FL CYTOKINE RECEPTOR P	9.64e+00
30	51	65.4	992	1	Y025_MYCPN	HYPOTHETICAL PROTEIN M	1.51e+01
31	50	64.1	299	1	VMAT_MEASY	MATRIX PROTEIN	1.51e+01
32	50	64.1	300	1	VEY2_YEAST	HYPOTHETICAL 49.5 KD P	1.51e+01
33	50	64.1	443	1	DHAC_CHICK	ALDEHYDE DEHYDROGENASE	1.51e+01
34	50	64.1	509	1	Y390_MYCPN	HYPOTHETICAL ATP-BINDI	1.51e+01
35	50	64.1	660	1	LCF1_YEAST	LONG-CHAIN-FATTY-ACID-LCF1_YEAST	1.51e+01
36	50	64.1	700	1	DSC2_BOVIN	DESMOCOLLIN 2A/2B PREC	1.51e+01
37	50	64.1	863	1	DSC2_HUMAN	DESMOCOLLIN 2A/2B PREC	1.51e+01
38	50	64.1	901	1	DSC2_MOUSE	DESMOCOLLIN 2A/2B PREC	1.51e+01
39	50	64.1	902	1	Y228_BORBU	HYPOTHETICAL PROTEIN B	1.51e+01
40	50	64.1	971	1	R115_YEAST	SERINE/THREONINE-PROTE	1.51e+01
41	50	64.1	1770	1	MYO5_DICDI	MYOSIN IJ HEAVY CHAIN	1.51e+01
42	50	64.1	2245	1	YR05_CABEL	HYPOTHETICAL 30.7 KD P	2.36e+01
43	49	62.8	269	1	FMT_SYNY3	METHIONYL-TRNA FORMYLT	2.36e+01
44	49	62.8	330	1	ADH2_KLULA	ALCOHOL DEHYDROGENASE	2.36e+01
45	49	62.8	348	1	ADH2_KLULA	ALCOHOL DEHYDROGENASE	2.36e+01

ALIGNMENTS

RESULT 1 STANDARD; PRT; 314 AA.  
AC Q64682;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DE 01-NOV-1997 (Rel. 35, Last annotation update)  
DE CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).  
GN TP53.  
OS Spermophilus beecheyi (Beechey ground squirrel).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Rodentia; Sciurognathi; Sciuridae; Scuriinae; Spermophilus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=THYMUS;  
RX MEDLINE; 95007566.  
RA RIVKINA M.B., CULLEN J.M., ROBINSON W.S., MARION P.L.;  
RT "State of the p53 gene in hepatocellular carcinomas of ground  
RT squirrels and woodchucks with past and ongoing infection with  
RT hepadnaviruses."  
RL Cancer Res. 54:5430-5437(1994).  
CC -1- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES  
CC GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL  
CC CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN  
CC TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION, IT IS A  
CC TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION  
CC BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF  
CC THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.  
CC APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF  
CC BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2  
CC EXPRESSION.  
CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
CC -1- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY  
CC OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED  
CC IN MANY TYPES OF CANCER.  
CC -1- SIMILARITY: BELONGS TO THE P53 FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; U43902; AAA85628.1;  
CC HSSP; P04637; LYCS.  
CC PROSITE; PS00348; P53; 1.



THIS PAGE BLANK (USPTO)

Job time : 17 secs.

```
RESULT 14
ENTRY
  TITLE      #type complete
  ORGANISM   nucleotide pyrophosphatase homolog - rice
  DATE       #formal_name Oryza sativa #common_name rice
  24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change
  24-Mar-1999
ACCESSIONS
  T03293
REFERENCE
  Z14889
  Hsing, Y.C.; Tsao, C.V.; Chow, T.; Hsieh, J.; Chen, Z.
  #authors   submitted to the EMBL Data Library, April 1995
  #submission Rice early embryogenesis gene.
  #description
  #accession  T03293
  ##status   preliminary; translated from GB/EMBL/DBJ
  ##molecule_type mRNA
  ##residues 1-479 #label HSI
  ##cross-references EMBL:U25430; NID:g818848; PID:g818849
  ##experimental_source strain Taining 67
GENETICS
  #note
SUMMARY
  OSE4
  #length 479 #molecular-weight 52198 #checksum 7353
  Query Match      69.2%; Score 54; DB 2; Length 479;
  Best Local Similarity 55.6%; Pred. No. 5.83e+00;
  Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
Db 161 PISGDYFTM 169
I:||||:
QY 2 PLDGEYFTL 10

RESULT 15
ENTRY
  TITLE      #type complete
  ORGANISM   conserved hypothetical protein slr0551 - Synechocystis sp.
  DATE       #strain PCC 6803)
  25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change
  24-Oct-1998
ACCESSIONS
  S76024
REFERENCE
  S74322
  Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.;
  Nakamura, Y.; Miyajima, N.; Hirose, M.; Sugita, M.;
  Sasamoto, S.; Kimura, T.; Hosouchi, T.; Matsuno, A.;
  Muraki, A.; Nakazaki, N.; Naruo, K.; Okumura, S.; Shimpō,
  S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.;
  Yasuda, M.; Tabata, S.
  #journal     DNA Res. (1996) 3:109-136
  #title       Sequence analysis of the genome of the unicellular
  cyanobacterium Synechocystis sp. PCC6803. II. Sequence
  determination of the entire genome and assignment of
  potential protein-coding regions.
  #cross-references MIM:37061201
  #accession   S76024
  ##molecule_type DNA
  ##residues 1-640 #label KAN
  ##cross-references EMBL:D64006; GB:AB001339; NID:gi001291; PID:d1011522;
  PID:gl001381
  ##note       the nucleotide sequence was submitted to the EMBL Data
  Library, June 1996
CLASSIFICATION
  #superfamily conserved hypothetical protein MGL39
SUMMARY
  #length 640 #molecular-weight 70427 #checksum 4335
  Query Match      69.2%; Score 54; DB 2; Length 640;
  Best Local Similarity 66.7%; Pred. No. 5.83e+00;
  Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Db 174 PIDGEFFDL 182
I:||||:
QY 2 PLDGEYFTL 10
```

Search completed: Sat Apr 15 01:05:36 2000

```

Db 324 KPLDGEYFTL 333
QY 1 KPLDGEYFTL 10

RESULT 10
ENTRY S51428 #type complete
TITLE hypothetical protein YLR183c - yeast (Saccharomyces cerevisiae)
ALTERNATE_NAMES hypothetical protein L9470.22
ORGANISM #formal_name Saccharomyces cerevisiae
DATE 23-Feb-1995 #sequence_revision 12-May-1995 #text_change 12-Dec-1997
ACCESSIONS S51428
REFERENCE S51414
#authors Wohldmann, P.
#submission submitted to the EMBL Data Library, November 1994
#description The sequence of S. cerevisiae cosmid 9470.
#accession S51428
##molecule_type DNA
##residues 1-489 ##label WOH
##cross-references EMBL:U17246; NID:g577192; PID:g577214; MIPS:YLR183c
GENETICS
#map_position 12R
SUMMARY #length 489 #molecular-weight 55467 #checksum 8282
Query Match 73.1%; Score 57; DB 2; Length 489;
Best Local Similarity 87.5%; Pred. No. 1.53e+00;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 440 KPLDEEYF 447
QY 1 KPLDGEYF 8

RESULT 11
ENTRY S74224 #type complete
TITLE aldehyde dehydrogenase (NAD+) (EC 1.2.1.3) 2 - mouse
ALTERNATE_NAMES retinaldehyde-specific dehydrogenase
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE 04-Dec-1997 #sequence_revision 12-Dec-1997 #text_change 17-Mar-1999
ACCESSIONS S74224
REFERENCE S74224
#authors Zhao, D.; McCaffery, P.; Ivins, K.J.; Neve, R.L.; Hogan, P.; Chin, W.W.; Draeger, U.C.
#journal Eur. J. Biochem. (1996) 240:15-22
#title Molecular identification of a major retinoic-acid-synthesizing enzyme, a retinaldehyde-specific dehydrogenase.
#cross-references MUID:96390857
#accession S74224
##molecule_type mRNA
##residues 1-499 ##label ZHA
##cross-references EMBL:X9273; NID:g1430868; PID:e254167; PID:g1430869
##experimental_source strain C3H/He; cell type embryonal carcinoma; cell line P19 teratocarcinoma induced with retinoic acid
GENETICS
#gene RALDH-2
CLASSIFICATION #superfamily aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology
KEYWORDS NAD; oxidoreductase
FEATURE
57-321 #domain aldehyde dehydrogenase homology #label ALDH
193-273 #domain NAD binding #status predicted #label NAD
267,301 #active_site Glu, Cys #status predicted
454 #binding_site NAD (Cys) #status predicted
SUMMARY #length 499 #molecular-weight 54725 #checksum 4300
Query Match 71.8%; Score 56; DB 2; Length 499;
Best Local Similarity 66.7%; Pred. No. 2.40e+00;

Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 144 PVDGYFTF 152
QY 2 PLDGEYFTL 10

RESULT 12
ENTRY S49113 #type complete
TITLE hypothetical protein 2 - Microcystis aeruginosa
ORGANISM #formal_name Microcystis aeruginosa
DATE 01-Feb-1995 #sequence_revision 12-May-1995 #text_change 09-Sep-1997
ACCESSIONS S49113
REFERENCE S49111
#authors Juerchott, K.; Boerner, T.
#submission submitted to the EMBL Data Library, November 1993
#description Sequence of the cyanobacterial plasmid pWAL from Microcystis aeruginosa HUB 5-2-4.
#accession S49113
##status preliminary
##molecule_type DNA
##residues 1-502 ##label JUE
##cross-references EMBL:Z28337; NID:g509352; PID:g509354
SUMMARY #length 502 #molecular-weight 58859 #checksum 5514
Query Match 71.8%; Score 56; DB 2; Length 502;
Best Local Similarity 77.8%; Pred. No. 2.40e+00;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 283 KTLGGEYFT 291
QY 1 KPLDGEYFT 9

RESULT 13
ENTRY F71884 #type complete
TITLE probable lipopolysaccharide biosynthesis protein - Helicobacter pylori (strain J99)
ORGANISM #formal_name Helicobacter pylori
#variety strain J99
DATE 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 12-Feb-1999
ACCESSIONS F71884
REFERENCE A71800
#authors Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.; Noonan, B.; Guild, B.C.; deJonge, B.L.; Carmel, G.; Tummino, P.J.; Caruso, A.; Uria-Nickelsen, M.; Mills, D.M.; Ives, C.; Gibson, R.; Marberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.; Trust, T.J.
#journal Nature (1999) 397:176-180
#title Genomic sequence comparison of two unrelated isolates of the human gastric pathogen Helicobacter pylori.
#cross-references MUID:99120557
#accession F71884
##status preliminary
##molecule_type DNA
##residues 1-373 ##label ARN
##cross-references GB:AE001511; GB:AE001439; NID:g4155382; PID:g4155383
##experimental_source strain J99
GENETICS
#gene jhp0820
SUMMARY #length 373 #molecular-weight 43230 #checksum 9679
Query Match 69.2%; Score 54; DB 2; Length 373;
Best Local Similarity 85.7%; Pred. No. 5.83e+00;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 154 PMDGEYF 160
QY 2 PLDGEYF 8

```

```
##status translated from GB/EMBL/DBJ
##molecule_type mRNA
##residues 1-212,'Q',214-393 ##label F10
##cross-references EMBL:X60019; NID:g506450; PID:g506451
#accession I38092
##status translated from GB/EMBL/DBJ
##molecule_type mRNA
##residues 1-253,'D',255-393 ##label F11
##cross-references EMBL:X60020; NID:g506452; PID:g506453
##note all sequences submitted to the EMBL/GenBank/DBJ
databases June 1991
REFERENCE
I38093
#authors Futreal, P.A.; Barrett, J.C.; Wiseman, R.W.
#journal Nucleic Acids Res. (1991) 19:6977
#title An Alu polymorphism intragenic to the TP53 gene.
#cross-references MUID:92107726
#accession I38093
##status translated from GB/EMBL/DBJ
##molecule_type DNA
##residues 1-393 ##label FUT
##cross-references EMBL:X54156; NID:g35213; PID:g35214
A44905
#authors Yamada, Y.; Yoshida, T.; Hayashi, K.; Sekiya, T.; Yokota, J.;
Hirohashi, S.; Nakatani, K.; Nakano, H.; Sugimura, T.;
Terada, M.
#journal Cancer Res. (1991) 51:5800-5805
#title p53 gene mutations in gastric cancer metastases and in
gastric cancer cell lines derived from metastases.
#cross-references MUID:92034678
#accession A44905
...
Note: remainder of annotations omitted.
Query Match 100.0%; Score 78; DB 1; Length 393;
Best Local Similarity 100.0%; Pred. No. 4.43e-05;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 321 KPLDGEYFTL 330
| | | | | | | | | |
QY 1 KPLDGEYFTL 10

RESULT 7
ENTRY JC6193 #type complete
TITLE tumor suppressor p53 - rabbit
ORGANISM #formal_name Oryctolagus cuniculus #common_name domestic
rabbit
DATE 11-Apr-1997 #sequence_revision 09-May-1997 #text_change
17-Mar-1999
ACCESSION JC6193
REFERENCE JC6193
#authors Le Goas, F.; May, P.; Ronco, P.; de Fromental, C.C.
#journal Gene (1997) 185:169-173
#title cDNA cloning and immunological characterization of rabbit
p53.
#cross-references MUID:97208869
#accession JC6193
##molecule_type mRNA
##residues 1-391 ##label LEA
##cross-references EMBL:X90592; NID:gl532043; PID:el94962; PID:gl532044
GENETICS
#gene p53
#accession JC6193
#superfamily cellular tumor antigen p53
CLASSIFICATION
#tumor
KEYWORDS #length 391 #molecular-weight 43435 #checksum 4367
SUMMARY
Query Match 93.6%; Score 73; DB 2; Length 391;
Best Local Similarity 90.0%; Pred. No. 6.15e-04;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 319 KPLDGEYFTL 328
| | | | | | | | | |
QY 1 KPLDGEYFTL 10

##status translated from GB/EMBL/DBJ
##molecule_type mRNA
##residues 1-212,'Q',214-393 ##label F10
##cross-references EMBL:X60019; NID:g506450; PID:g506451
#accession I38092
##status translated from GB/EMBL/DBJ
##molecule_type mRNA
##residues 1-253,'D',255-393 ##label F11
##cross-references EMBL:X60020; NID:g506452; PID:g506453
##note all sequences submitted to the EMBL/GenBank/DBJ
databases June 1991
REFERENCE
I38093
#authors Futreal, P.A.; Barrett, J.C.; Wiseman, R.W.
#journal Nucleic Acids Res. (1991) 19:6977
#title An Alu polymorphism intragenic to the TP53 gene.
#cross-references MUID:92107726
#accession I38093
##status translated from GB/EMBL/DBJ
##molecule_type DNA
##residues 1-393 ##label FUT
##cross-references EMBL:X54156; NID:g35213; PID:g35214
A44905
#authors Yamada, Y.; Yoshida, T.; Hayashi, K.; Sekiya, T.; Yokota, J.;
Hirohashi, S.; Nakatani, K.; Nakano, H.; Sugimura, T.;
Terada, M.
#journal Cancer Res. (1991) 51:5800-5805
#title p53 gene mutations in gastric cancer metastases and in
gastric cancer cell lines derived from metastases.
#cross-references MUID:92034678
#accession A44905
...
Note: remainder of annotations omitted.
Query Match 100.0%; Score 78; DB 1; Length 393;
Best Local Similarity 100.0%; Pred. No. 4.43e-05;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 321 KPLDGEYFTL 330
| | | | | | | | | |
QY 1 KPLDGEYFTL 10

RESULT 9
ENTRY JH0633 #type complete
TITLE cellular tumor antigen p53 - golden hamster
ALTERNATE_NAMES tumor-suppressor protein p53
ORGANISM #formal_name Mesocricetus auratus #common_name golden hamster
DATE 17-Aug-1992 #sequence_revision 17-Aug-1992 #text_change
08-Sep-1997
ACCESSION JH0633
REFERENCE JH0633
#authors Legros, Y.; McIntyre, P.; Soussi, T.
#journal Gene (1992) 112:247-250
#title The cDNA cloning and immunological characterization of
hamster p53.
#cross-references MUID:92210007
#accession JH0633
##molecule_type mRNA
##residues 1-396 ##label LEG
##cross-references GB:M75144; NID:gl91414; PID:gl91415
##experimental_source kidney, strain MPI
GENETICS
#gene p53
#accession JH0633
#superfamily cellular tumor antigen p53
CLASSIFICATION
#apoptosis; cell division control; DNA binding; homotetramer;
nucleus; phosphoprotein; transcription regulation; tumor
suppressor; zinc
KEYWORDS #binding_site predicted\
#binding_site phosphoryl-RNA (Ser) (covalent) #status
predicted
FEATURE 179,182,241,245 #binding_site zinc (Cys, His, Cys, Cys) #status
395 #length 396 #molecular-weight 43631 #checksum 6617
SUMMARY
Query Match 89.7%; Score 70; DB 2; Length 396;
Best Local Similarity 90.0%; Pred. No. 2.87e-03;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```



```
16-26 #region conserved region I\
99-289 #domain DNA-binding core #status predicted #label DBC\
108-121 #region L1 loop\
114-139 #region conserved region II\
160-192 #region L2 loop\
168-178 #region conserved region III\
231-252 #region conserved region IV\
233-248 #region L3 loop\
267-283 #region conserved region V\
313-319 #region nuclear location signal\
319-357 #region tetramer association\
7,9,12,18,23,37 #binding_site phosphate (Ser) (covalent) #status
173,176,235,239 #binding_site zinc (Cys, His, Cys, Cys) #status
312 #binding_site phosphate (Ser) (covalent) (by cdc2
389 #binding_site phosphoryl-RNA (Ser) (covalent) #status
SUMMARY #length 390 #molecular-weight 43458 #checksum 1260
Query Match 100.0%; Score 78; DB 1; Length 390;
Best Local Similarity 100.0%; Pred. No. 4.43e-05;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 318 KPLDGEYFTL 327
|||||
QY 1 KPLDGEYFTL 10

RESULT 4
ENTRY S02192 #type complete
TITLE cellular tumor antigen p53 - rat
ALTERNATE_NAMES gene p53 protein; nuclear oncoprotein p53
ORGANISM #formal_name Rattus norvegicus #common_name Norway rat
DATE 18-Oct-1989 #sequence_revision 18-Oct-1989 #text_change
17-Mar-1999
ACCESSIONS S02192; S41149
REFERENCE S02192
#authors Soussi, T.; de Fromental, C.C.; Breugnot, C.; May, E.
#journal Nucleic Acids Res. (1988) 16:11384
#title Nucleotide sequence of a cDNA encoding the rat p53 nuclear
oncoprotein.
#cross-references MUID:89083585
#accession S02192
#molecule_type mRNA
#residues 1-391 #label SOU
#cross-references EMBL:X13058; NID:g56828; PID:g56829
REFERENCE S41149
#authors Hulla, J.E.; Schneider, R.P.
#journal Nucleic Acids Res. (1993) 21:713-717
#title Structure of the rat p53 tumor suppressor gene.
#cross-references MUID:93181268
#accession S41149
#status preliminary; nucleic acid sequence not shown;
translation not shown
#molecule_type DNA
#residues 1-173, 'W', 175-391 #label HUL
#cross-references EMBL:L07909
#note the nucleotide sequence was submitted to the EMBL Data
Library, December 1992
GENETICS
#introns 25/2; 32/3; 185/1; 259/2; 305/1; 329/3; 365/2
CLASSIFICATION #superfamily cellular tumor antigen p53
KEYWORDS apoptosis; cell division control; DNA binding; homotetramer;
nucleus; phosphoprotein; transcription regulation; tumor
suppressor; zinc
FEATURE
174,177,236,240 #binding_site zinc (Cys, His, Cys, Cys) #status
390 #binding_site phosphoryl-RNA (Ser) (covalent) #status
predicted
SUMMARY #length 391 #molecular-weight 43451 #checksum 7105
```

```
Query Match 100.0%; Score 78; DB 2; Length 391;
Best Local Similarity 100.0%; Pred. No. 4.43e-05;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 319 KPLDGEYFTL 328
|||||
QY 1 KPLDGEYFTL 10

RESULT 5
ENTRY S06594 #type complete
TITLE cellular tumor antigen p53 - green monkey
ORGANISM #formal_name Cercopithecus aethiops #common_name green
monkey, grivet
DATE 28-Feb-1990 #sequence_revision 28-Feb-1990 #text_change
08-Sep-1997
ACCESSIONS S06594
REFERENCE S06594
#authors Rigaudy, P.; Eckhart, W.
#journal Nucleic Acids Res. (1989) 17:8375
#title Nucleotide sequence of a cDNA encoding the monkey cellular
phosphoprotein p53.
#cross-references MUID:90045967
#accession S06594
#molecule_type mRNA
#residues 1-393 #label RIG
#cross-references EMBL:X16384; NID:g22795; PID:g22796
CLASSIFICATION #superfamily cellular tumor antigen p53
KEYWORDS apoptosis; cell division control; DNA binding; homotetramer;
nucleus; phosphoprotein; transcription regulation; tumor
suppressor; zinc
FEATURE
176,179,238,242 #binding_site zinc (Cys, His, Cys, Cys) #status
392 #binding_site phosphoryl-RNA (Ser) (covalent) #status
predicted
SUMMARY #length 393 #molecular-weight 43696 #checksum 4263
Query Match 100.0%; Score 78; DB 2; Length 393;
Best Local Similarity 100.0%; Pred. No. 4.43e-05;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 321 KPLDGEYFTL 330
|||||
QY 1 KPLDGEYFTL 10

RESULT 6
ENTRY DNHU53 #type complete
TITLE cellular tumor antigen p53 - human
ALTERNATE_NAMES cellular phosphoprotein p53; oncoprotein p53; transformation
suppressor p53; tumor suppressor p53
ORGANISM #formal_name Homo sapiens #common_name man
DATE 05-Oct-1988 #sequence_revision 18-Nov-1994 #text_change
26-Feb-1999
ACCESSIONS A25224; A43073; J04036; S40773; S42669; A22837; A55060;
A25397; B25397; S42452; S42453; I38082; I38083; I38084;
I38085; I38086; I38087; I38088; I38089; I38090; I38091;
I38092; I38093; A44905; I58354; I78850; I52861; S60153
REFERENCE A25224
#authors Lamb, P.; Crawford, L.
#journal Mol. Cell. Biol. (1986) 6:1379-1385
#title Characterization of the human p53 gene.
#cross-references MUID:87064416
#accession A25224
#molecule_type DNA
#residues 1-393 #label LAM
#cross-references EMBL:X01405; GB:MI3121; GB:N00032; NID:g189460;
PID:g386994
REFERENCE J04036
#authors Buchman, V.L.; Chumakov, P.M.; Ninkina, N.N.; Samarina, O.P.;
Georgiev, G.P.
```

```

233-248      #region L3 loop\
267-283      #region conserved region V\
313-319      #region nuclear location signal\
319-357      #region tetramer association\
7,9,12,18,23,37 #binding_site phosphate (Ser) (covalent) #status
173,176,235,239 #binding_site zinc (Cys, His, Cys, Cys) #status
312          #binding_site phosphate (Ser) (covalent) (by cdc2
              kinase) #status predicted
SUMMARY      #length 381 #molecular_weight 42498 #checksum 8703
Query Match  100.0%; Score 78; DB 2; Length 381;
Best Local Similarity 100.0%; Pred. NO. 4.43e-05;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 318 KPLDGEYFTL 327
QY 1 KPLDGEYFTL 10

RESULT 2
ENTRY      S51648      #type complete
TITLE      cellular tumor antigen p53 - bovine
ALTERNATE_NAMES tumor-suppressor protein p53
ORGANISM   #formal_name Bos primigenius taurus #common_name cattle
DATE       07-May-1995 #sequence_revision 01-Sep-1995 #text_change
08-Sep-1997
ACCESSIONS S51648
REFERENCE   S51648
#authors   Dequiedt, F.; Willems, L.; Burny, A.; Kettmann, R.
#submission submitted to the EMBL data library, September 1994
#description Nucleotide sequence of the ovine p53 tumor-suppressor gene
              cDNA and its genomic organisation.
#accession S51648
#status    preliminary
#molecule_type mRNA
#residues  1-386 #label DEQ
#cross-references EMBL:X81704; NID:g602332; PID:g602333
CLASSIFICATION #superfamily cellular tumor antigen p53
KEYWORDS      apoptosis; cell division control; DNA binding; homotetramer;
              phosphoprotein; transcription regulation; tumor suppressor;
              zinc
FEATURE
168,171,231,235 #binding_site zinc (Cys, His, Cys, Cys) #status
385             #binding_site phosphoryl-RNA (Ser) (covalent) #status
SUMMARY      #length 386 #molecular_weight 43255 #checksum 7025
Query Match  100.0%; Score 78; DB 2; Length 386;
Best Local Similarity 100.0%; Pred. NO. 4.43e-05;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 314 KPLDGEYFTL 323
QY 1 KPLDGEYFTL 10

RESULT 3
ENTRY      DNMS53      #type complete
TITLE      cellular tumor antigen p53 - mouse
ALTERNATE_NAMES oncoprotein p53
ORGANISM   #formal_name Mus musculus #common_name house mouse
DATE       28-Aug-1985 #sequence_revision 04-Oct-1996 #text_change
12-Feb-1999
ACCESSIONS A22739; S06336; A02684; S38822; S38823; S40014; I48703
REFERENCE   A22739
#authors   Bienz, B.; Zakut-Houri, R.; Givol, D.; Oren, M.
#journal   EMBO J. (1984) 3:2179-2183
#cross-references MUID:8502173
#accession A22739
#molecule_type DNA

```

```

#residues      1-134,'V',136-390 #label BIE
#cross-references GB:X00876; NID:g871420; PID:g871421; GB:X01237;
              GB:K01700; NID:g53575; PID:g53576
REFERENCE      S06336
#authors       Chumakov, P.M.
#journal       Bioorg. Khim. (1987) 13:1691-1694
#title         Primary structure of DNA complementary to murine oncoprotein
              p53 mRNA.
#cross-references MUID:88221682
#accession     S06336
#status        not compared with conceptual translation
#molecule_type mRNA
#residues      1-134,'V',136-390 #label CHU
REFERENCE      A02684
#authors       Zakut-Houri, R.; Oren, M.; Bienz, B.; Lavie, V.; Hazum, S.;
              Givol, D.
#journal       Nature (1983) 306:594-597
#title         A single gene and a pseudogene for the cellular tumour
              antigen p53.
#cross-references MUID:84068204
#accession     A02684
#molecule_type mRNA
#residues      1-159,'H',161-167,'G',169-233,'I',235-390 #label ZAK
#cross-references GB:X01237; GB:K01700; NID:g53575
REFERENCE      S38822
#authors       Arai, N.; Nomura, D.; Yokota, K.; Wolf, D.; Brill, E.;
              Shohat, O.; Rotter, V.
#journal       Mol. Cell. Biol. (1986) 6:3232-3239
#title         Immunologically distinct p53 molecules generated by
              alternative splicing.
#cross-references MUID:87064640
#accession     S38822
#status        preliminary
#molecule_type mRNA
#residues      1-390 #label ARA1
#cross-references EMBL:M13872; NID:g200198; PID:g200199
#accession     S38823
#status        preliminary
#molecule_type mRNA
#residues      1-167,'G',169-233,'I',235-390 #label ARA2
#cross-references EMBL:M13873
REFERENCE      S40014
#authors       Arai, N.; Nomura, D.; Yokota, K.; Wolf, D.; Brill, E.;
              Shohat, O.; Rotter, V.
#submission    submitted to the EMBL Data Library, July 1988
#accession     S40014
#molecule_type mRNA
#residues      1-167,'G',169-390 #label ARA3
#cross-references EMBL:M13873; NID:g200200; PID:g200201
REFERENCE      I48703
#authors       Jenkins, J.R.; Rudge, K.; Redmond, S.; Wade-Evans, A.
#journal       Nucleic Acids Res. (1984) 12:5609-5626
#title         Cloning and expression analysis of full length mouse cDNA
              sequences encoding the transformation associated protein
              p53
#cross-references MUID:84272240
#accession     I48703
#status        preliminary; translated from GB/EMBL/DBDJ
#molecule_type mRNA
#residues      1-47,'R',49-78,'QW',82-390 #label RES
#cross-references EMBL:X00741; NID:g53570; PID:g53571
COMMENT       This DNA-binding protein plays an essential role in the regulation
              of cell division, as it is required for the transition from phase
              G0 to G1 of the cell cycle.
COMMENT       The tetramer association region may exhibit a beta-turn,
              beta-sheet, beta-turn, alpha-helix motif.
CLASSIFICATION #superfamily cellular tumor antigen p53
KEYWORDS      apoptosis; cell division control; DNA binding; homotetramer;
              phosphoprotein; transcription regulation; tumor suppressor;
              zinc
FEATURE
1-44          #domain transcription activation #status predicted
              #label TRA\

```

(TT)

Result No.	Score	Query		Length	DB	ID	Description	Pred. No.
		Match						
1	78	100.0	381	2	S38824		cellular tumor antigen	4.43e-05
2	78	100.0	386	2	S516A8		cellular tumor antigen	4.43e-05
3	78	100.0	390	1	DNMS53		cellular tumor antigen	4.43e-05
4	78	100.0	391	2	S02192		cellular tumor antigen	4.43e-05
5	78	100.0	393	2	S06594		cellular tumor antigen	4.43e-05
6	78	100.0	393	1	DNHU53		cellular tumor antigen	4.43e-05
7	73	93.6	391	2	JC6193		tumor suppressor p53	6.15e-04
8	70	89.7	393	2	JC6176		tumor suppressor prot	2.87e-03
9	70	89.7	396	2	JH0633		cellular tumor antigen	2.87e-03
10	57	73.1	489	2	S51428		hypothetical protein	1.53e+00
11	56	71.8	499	2	S74224		aldehyde dehydrogenas	2.40e+00
12	56	71.8	502	2	S49113		hypothetical protein	2.40e+00
13	54	69.2	373	2	F71884		probable lipopolysacc	5.83e+00
14	54	69.2	479	2	T03293		nucleotide pyrophosph	5.83e+00
15	54	69.2	640	2	S76024		conserved hypothetical	5.83e+00
16	54	69.2	694	2	B54901		long-chain-fatty-acid	5.83e+00
17	54	69.2	694	2	S56060		long-chain-fatty-acid	5.83e+00
18	53	67.9	183	2	S05358		hypothetical protein	9.01e+00
19	53	67.9	238	2	S76860		hypothetical protein	9.01e+00
20	53	67.9	305	2	P64481		hypothetical protein	9.01e+00
21	53	67.9	520	1	S09030		aldehyde dehydrogenas	9.01e+00
22	53	67.9	521	2	S09599		hypothetical protein	9.01e+00
23	52	66.7	237	2	D71182		hypothetical protein	1.38e+01



PI Bracco L, Conseiller E;  
 DR WPI; 97-132633/12.  
 DR N-PSDB; T86216.  
 PT New p53 variants e.g. with oligomerisation domain replaced by  
 PT leucine zipper - useful for treating hyper-proliferative disorders,  
 PT esp. cancer and restenosis  
 PS Claim 31; Pages 78-80; 13pp; French.  
 CC Claimed variants of protein p53 have at least part of the  
 CC oligomerisation domain deleted and replaced by a leucine zipper  
 CC domain. The mutants preferably also have at least part of the p53  
 CC transactivation domain (amino acids 1-74) deleted and replaced by  
 CC the transactivating domain (TD) from herpes simplex virus viral  
 CC protein VP16 (amino acids 411-490). The present sequence is that of  
 CC a specifically claimed p53 variant designated V-336 and comprising  
 CC the VP16 TD, amino acids 75-336 of human wild-type p53 and a  
 CC leucine zipper domain at the C-terminal. The p53 variants are  
 CC more active and more stable tumour suppressors and apoptosis-inducing  
 CC agents than wild-type p53 and are active where the wild-type protein  
 CC is not, i.e. they are not inactivated by dominant negative or oncogenic  
 CC mutants, nor by other cellular proteins (because the leucine zipper  
 CC domain prevents formation of inactive mixed oligomers).  
 SQ Sequence 374 AA;

Query Match 100.0%; Score 78; DB 1; Length 374;  
 Best Local Similarity 100.0%; Pred. No. 6.27e-02;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 329 KPLDGEYFTL 338

QY 1 KPLDGEYFTL 10  
 |||||

Search completed: Sat Apr 15 01:05:00 2000  
 Job time : 35 secs.

```

RESULT 14
ID W28482 standard; Protein; 374 AA.
AC W28482;
DT 25-NOV-1997 (first entry)
DE Human p53 protein variant V-336H.
KW Leucine zipper domain; LZD; Oligomerisation domain; mutant; mutein;
KW substitution; replacement; transactivation; viral protein VP16; HSV;
KW anti-oncogene; hyperproliferation; cancer; restenosis;
KW tumour suppression; apoptosis.
OS Chimeric - Homo sapiens.
OS Chimeric - Homo sapiens.
OS Synthetic.
OS Key Location/Qualifiers
FT misc_difference 189
FT /note= "Arg residue at position 182 of wild-type
FT p53 has been mutated to His"
PN WO9704092-A1.
PD 06-FEB-1997.
PD 17-JUL-1996; F01111.
PD 19-JUL-1995; FR-008729.
PR (RHON ) RHONE POULENC RORER SA.
PI Bracco, L; Conseiller E;
PI WPI; 97-132633/12.
DR New p53 variants e.g. with oligomerisation domain replaced by
PT leucine zipper - useful for treating hyper-proliferative disorders,
PT esp. cancer and restenosis
PT Claim 31; Page -; 133pp; French.
CC Claimed variants of protein p53 have at least part of the
CC oligomerisation domain deleted and replaced by a leucine zipper
CC domain. The mutants preferably also have at least part of the p53
CC transactivation domain (amino acids 1-74) deleted and replaced by
CC the transactivating domain (TD) from herpes simplex virus viral
CC protein VP16 (amino acids 411-490). The present sequence is that of
CC a specifically claimed p53 variant designated V-336H and comprising
CC the VP16 TD, amino acids 75-336 of human wild-type p53 (but with
CC Arg182 replaced by His) and a leucine zipper domain at the C-terminal.
CC The p53 variants are more active and more stable tumour suppressors
CC and apoptosis-inducing agents than wild-type p53 and are active where
CC the wild-type protein is not, i.e. they are not inactivated by dominant
CC negative or oncogenic mutants, nor by other cellular proteins (because
CC the leucine zipper domain prevents formation of inactive mixed
CC oligomers).
CC (Note: this sequence does not appear in the specification and has
CC been produced by modifying the given sequence of variant V-336).
CC Sequence 374 AA;
SQ
Query Match 100.0%; Score 78; DB 1; Length 374;
Best Local Similarity 100.0%; Pred. NO. 6,27e-02;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 329 KPLDGEYFTL 338
QY 1 KPLDGEYFTL 10
RESULT 15
ID W28481 standard; Protein; 374 AA.
AC W28481;
DT 25-NOV-1997 (first entry)
DE Human p53 protein variant V-336 encoded by pEC116.
KW Leucine zipper domain; LZD; Oligomerisation domain; mutant; mutein;
KW substitution; replacement; transactivation; viral protein VP16; HSV;
KW anti-oncogene; hyperproliferation; cancer; restenosis;
KW tumour suppression; apoptosis.
OS Chimeric - Homo sapiens.
OS Chimeric - Herpes simplex virus.
OS Synthetic.
OS Key Location/Qualifiers
FT misc_difference 189
FT /note= "Arg residue at position 182 of wild-type
FT p53 has been mutated to His"
PN WO9704092-A1.
PD 06-FEB-1997.
PD 17-JUL-1996; F01111.
PD 19-JUL-1995; FR-008729.
PR (RHON ) RHONE POULENC RORER SA.

```

CC p53 constructs (see also W13954, W13956-61, W13971-77) bearing  
CC a deletion of all or a fragment of the C-terminal residues  
CC 356-393 have DNA binding ability and can activate the DNA binding  
CC of common Class I p53 tumour mutants (see also W13951-52). The  
CC method provides the means for pharmacological rescue of p53  
CC function in cancer patients. Nucleic acids coding for such  
CC constructs can be used for cancer gene therapy.  
SQ Sequence 355 AA;

Query Match 100.0%; Score 78; DB 1; Length 355;  
Best Local Similarity 100.0%; Pred. No. 6.27e-02;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 321 KPLDGEYFTL 330  
|||||  
QY 1 KPLDGEYFTL 10

RESULT 9  
ID W13975 standard; Protein; 363 AA.  
AC W13975;  
DT 25-JUN-1997 (first entry)  
DE Modified p53 variant p53H273R284del1364-393.  
KW p53; tumour suppressor; cancer; therapy; cell proliferation;  
KW apoptosis; protein engineering; DNA binding.  
OS Synthetic.  
PN W09710843-A1.  
PD 27-MAR-1997.  
PF 20-SEP-1996; U15188.  
PR 22-SEP-1995; US-004802.  
PR 21-AUG-1996; US-697221.  
PA (WIST-) WISTAR INST ANATOMY & BIOLOGY.  
PI Halazonetis TD;  
DR WPI: 97-202618/18.  
PT R284K modified p53 protein having DNA binding ability - useful in  
PT treatment of cancer  
FS Example 1; 58-59; 82pp; English.  
CC Modified p53 variant p53H273R284del1364-393 (W13975) has the tumour-  
CC derived His273 mutation (see also W13952), a Thr284 to Arg substn.  
CC (see also W13949) and a deletion of the 30 C-terminal amino acids  
CC of wild-type p53 (W13948). His273 is a Class I p53 tumour mutation  
CC that affects DNA binding. The T284R substitution, introduced by  
CC site-directed mutagenesis of p53 DNA, provides a novel p53-DNA  
CC contact between a phosphate of the DNA backbone and p53, and  
CC restores DNA binding. The C-terminal deletion permits in vitro  
CC DNA binding. The construct provides the means for pharmacological  
CC rescue of p53 function in cancer patients. Other modified p53  
CC constructs (W13949-50, W13953-54, W13968-77) have also been  
CC produced. Nucleic acids coding for modified p53 can be used for  
CC cancer gene therapy.  
SQ Sequence 363 AA;

Query Match 100.0%; Score 78; DB 1; Length 363;  
Best Local Similarity 100.0%; Pred. No. 6.27e-02;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 321 KPLDGEYFTL 330  
|||||  
QY 1 KPLDGEYFTL 10

RESULT 10  
ID W13972 standard; Protein; 363 AA.  
AC W13972;  
DT 25-JUN-1997 (first entry)  
DE Modified p53 variant p53Q248del1364-393.  
KW p53; tumour suppressor; cancer; therapy; cell proliferation;  
KW apoptosis; protein engineering; DNA binding.  
OS Synthetic.  
PN W09710843-A1.  
PD 27-MAR-1997.  
PF 20-SEP-1996; U15188.  
PR 22-SEP-1995; US-004802.

CC p53 constructs (see also W13954, W13956-61, W13971-77) bearing  
CC a deletion of all or a fragment of the C-terminal residues  
CC 356-393 have DNA binding ability and can activate the DNA binding  
CC of common Class I p53 tumour mutants (see also W13951-52). The  
CC method provides the means for pharmacological rescue of p53  
CC function in cancer patients. Nucleic acids coding for such  
CC constructs can be used for cancer gene therapy.  
SQ Sequence 355 AA;

Query Match 100.0%; Score 78; DB 1; Length 355;  
Best Local Similarity 100.0%; Pred. No. 6.27e-02;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 321 KPLDGEYFTL 330  
|||||  
QY 1 KPLDGEYFTL 10

RESULT 11  
ID W13954 standard; Protein; 363 AA.  
AC W13954;  
DT 25-JUN-1997 (first entry)  
DE Modified p53 variant (del1364-393).  
KW p53; tumour suppressor; cancer; therapy; cell proliferation;  
KW apoptosis; protein engineering; DNA binding.  
OS Synthetic.  
PN W09710843-A1.  
PD 27-MAR-1997.  
PF 20-SEP-1996; U15188.  
PR 22-SEP-1995; US-004802.  
PR 21-AUG-1996; US-697221.  
PA (WIST-) WISTAR INST ANATOMY & BIOLOGY.  
PI Halazonetis TD;  
DR WPI: 97-202618/18.  
PT R284K modified p53 protein having DNA binding ability - useful in  
PT treatment of cancer  
FS Example 1; 49-51; 82pp; English.  
CC A modified p53 variant (W13954) comprises wild-type p53 (see  
CC also W13948) having a deletion of the C-terminal 30 amino acids,  
CC and is obt'd. by site-directed mutagenesis of p53 DNA. Deletion of  
CC the p53 C-terminal 30 amino acids activates the DNA binding of  
CC common Class I p53 mutants (see also W13951-52). Novel modified  
CC p53 variants (W13949-50, W13953-54, W13968-77), some contg  
CC C-terminal deletions, provide the means for pharmacological rescue  
CC of p53 function in cancer patients. Nucleic acids coding for  
CC modified p53 can be used for cancer gene therapy.  
SQ Sequence 363 AA;

Query Match 100.0%; Score 78; DB 1; Length 363;  
Best Local Similarity 100.0%; Pred. No. 6.27e-02;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 321 KPLDGEYFTL 330  
|||||  
QY 1 KPLDGEYFTL 10

RESULT 12  
ID W13976 standard; Protein; 363 AA.  
AC W13976;  
DT 25-JUN-1997 (first entry)  
DE Modified p53 variant p53C273del1364-393.  
KW p53; tumour suppressor; cancer; therapy; cell proliferation;

```
Db 85 KPLDGEYFTL 94
QY 1 KPLDGEYFTL 10

RESULT 5
ID R51878 standard; Protein; 157 AA.
AC R51878;
DT 18-NOV-1994 (first entry)
DE Human p53 amino acids 237-393.
KW Human nuclear phosphoprotein p53; tumour suppressor gene product;
OS anti-oncogene; cancer; tumour; antibody binding region; epitope.
FH Key Location/Qualifiers
FT misc_difference 37 /note= "Arg corresponds to a CAT codon"
PN WO9408241-A.
PD 14-APR-1994.
PF 30-SEP-1993; E02666.
PR 30-SEP-1992; DE-232823.
PA (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM STIFTUNG.
PI Klein R, Schranz P, Tessmer C, Volkmann M, Zentgraf H;
DR WPI: 94-135732/16.
DR N-PSDB: Q62363.
PT Non-radioactive detection of p53 specific antibodies - by capture
PT on immobilised p53 or its fragments, then reaction with labelled
PT second antibody, for diagnosis of tumours and suitable for
PT screening
PS Claim 10; Page 19; 35pp; German.
CC Antibodies specific for p53 are detected by binding to immobilised
CC fragments of the p53 gene product containing the antibody-binding
CC region. Preferred fragments contain amino acids 1-241, 40-349,
CC 40-393, 66-241, 66-393, 237-349, 237-393 and esp. 9-33, 37-52 or
CC 368-386. See R51872-R51881 for sequences of these fragments.
SQ Sequence 157 AA;

Query Match 100.0%; Score 78; DB 1; Length 157;
Best Local Similarity 100.0%; Pred. No. 6.27e-02;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 85 KPLDGEYFTL 94
QY 1 KPLDGEYFTL 10

RESULT 6
ID R51873 standard; Protein; 310 AA.
AC R51873;
DT 18-NOV-1994 (first entry)
DE Human p53 amino acids 40-349.
KW Human nuclear phosphoprotein p53; tumour suppressor gene product;
OS anti-oncogene; cancer; tumour; antibody binding region; epitope.
FH Key Location/Qualifiers
FT misc_difference 234 /note= "Arg corresponds to a CAT codon"
PN WO9408241-A.
PD 14-APR-1994.
PF 30-SEP-1993; E02666.
PR 30-SEP-1992; DE-232823.
PA (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM STIFTUNG.
PI Klein R, Schranz P, Tessmer C, Volkmann M, Zentgraf H;
DR WPI: 94-135732/16.
DR N-PSDB: Q62358.
PT Non-radioactive detection of p53 specific antibodies - by capture
PT on immobilised p53 or its fragments, then reaction with labelled
PT second antibody, for diagnosis of tumours and suitable for
PT screening
PS Claim 10; Page 17; 35pp; German.
CC Antibodies specific for p53 are detected by binding to immobilised
CC fragments of the p53 gene product containing the antibody-binding
CC region. Preferred fragments contain amino acids 1-241, 40-349,
CC 40-393, 66-241, 66-393, 237-349, 237-393 and esp. 9-33, 37-52 or
```

```
CC 368-386. See R51872-R51881 for sequences of these fragments.
SQ Sequence 328 AA;

Query Match 100.0%; Score 78; DB 1; Length 328;
Best Local Similarity 100.0%; Pred. No. 6.27e-02;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 256 KPLDGEYFTL 265
QY 1 KPLDGEYFTL 10

RESULT 8
ID W13950 standard; Protein; 355 AA.
AC W13950;
DT 25-JUN-1997 (first entry)
DE Del356-393 modified human p53.
KW p53; tumour suppressor; cancer; therapy; cell proliferation;
OS apoptosis; protein engineering; DNA binding.
FH Key Location/Qualifiers
FT misc_difference 234 /note= "Arg corresponds to a CAT codon"
PN WO9710843-A1.
PD 27-MAR-1997.
PF 20-SEP-1996; U15188.
PR 22-SEP-1995; US-004802.
PR 21-AUG-1996; US-697221.
PA (WIST-) WISTAR INST ANATOMY & BIOLOGY.
PI Halazonetis TD;
DR WPI: 97-202618/18.
PT R284K modified p53 protein having DNA binding ability - useful in
PT treatment of cancer
PS Claim 3; Refer to Page 27-29; 82pp; English.
CC Del356-393 modified p53 (W13950) has the C-terminal region of
CC wild-type human p53 tumour suppressor (W13948) deleted. Modified
```

```
CC 368-386. See R51872-R51881 for sequences of these fragments.
SQ Sequence 310 AA;

Query Match 100.0%; Score 78; DB 1; Length 310;
Best Local Similarity 100.0%; Pred. No. 6.27e-02;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 282 KPLDGEYFTL 291
QY 1 KPLDGEYFTL 10

RESULT 7
ID R51876 standard; Protein; 328 AA.
AC R51876;
DT 18-NOV-1994 (first entry)
DE Human p53 amino acids 66-393.
KW Human nuclear phosphoprotein p53; tumour suppressor gene product;
OS anti-oncogene; cancer; tumour; antibody binding region; epitope.
FH Key Location/Qualifiers
FT misc_difference 208 /note= "Arg corresponds to a CAT codon"
PN WO9408241-A.
PD 14-APR-1994.
PF 30-SEP-1993; E02666.
PR 30-SEP-1992; DE-232823.
PA (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM STIFTUNG.
PI Klein R, Schranz P, Tessmer C, Volkmann M, Zentgraf H;
DR WPI: 94-135732/16.
DR N-PSDB: Q62361.
PT Non-radioactive detection of p53 specific antibodies - by capture
PT on immobilised p53 or its fragments, then reaction with labelled
PT second antibody, for diagnosis of tumours and suitable for
PT screening
PS Claim 10; Page 18; 35pp; German.
CC Antibodies specific for p53 are detected by binding to immobilised
CC fragments of the p53 gene product containing the antibody-binding
CC region. Preferred fragments contain amino acids 1-241, 40-349,
CC 40-393, 66-241, 66-393, 237-349, 237-393 and esp. 9-33, 37-52 or
CC 368-386. See R51872-R51881 for sequences of these fragments.
SQ Sequence 328 AA;

Query Match 100.0%; Score 78; DB 1; Length 328;
Best Local Similarity 100.0%; Pred. No. 6.27e-02;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 256 KPLDGEYFTL 265
QY 1 KPLDGEYFTL 10

RESULT 8
ID W13950 standard; Protein; 355 AA.
AC W13950;
DT 25-JUN-1997 (first entry)
DE Del356-393 modified human p53.
KW p53; tumour suppressor; cancer; therapy; cell proliferation;
OS apoptosis; protein engineering; DNA binding.
FH Key Location/Qualifiers
FT misc_difference 234 /note= "Arg corresponds to a CAT codon"
PN WO9710843-A1.
PD 27-MAR-1997.
PF 20-SEP-1996; U15188.
PR 22-SEP-1995; US-004802.
PR 21-AUG-1996; US-697221.
PA (WIST-) WISTAR INST ANATOMY & BIOLOGY.
PI Halazonetis TD;
DR WPI: 97-202618/18.
PT R284K modified p53 protein having DNA binding ability - useful in
PT treatment of cancer
PS Claim 3; Refer to Page 27-29; 82pp; English.
CC Del356-393 modified p53 (W13950) has the C-terminal region of
CC wild-type human p53 tumour suppressor (W13948) deleted. Modified
```

```

Query Match      100.0%; Score 78; DB 1; Length 48;
Best Local Similarity 100.0%; Pred. No. 6.27e-02;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 3 KPLDGEYFTL 12
QY 1 KPLDGEYFTL 10
|||||

RESULT 2
ID W22024 standard; Protein; 73 AA.
AC W22024;
DE Wild-type p53 tetramerising domain.
KW Globin analogue; GCN4; yeast transcription factor; oligomerising domain;
KW ligand binding domain; multimeric haemoglobin; oxygen carrier; anaemia;
KW blood substitute; therapy; haematopoesis; oxygen removal; Hb; p53;
KW nitric oxide removal; tetramerising domain.
OS Synthetic.
PN WO9723631-A2.
PD 03-JUL-1997.
PF 20-DEC-1996; U20632.
PR 22-DEC-1995; US-021001.
PA (SOMA-) SOMATOGEN INC.
PI Anthony-Cahill SJ, Epp JK, Kerwin BA, Mathews AJ;
PI Ollins PO;
DR WPI; 97-351067/32.
PT New globin containing non-natural binding site and related nucleic
PT acid - also multimeric haemoglobin, used as oxygen carrier for in
PT vivo or in vitro applications, with extended half-life and reduced
PT extravasation
PS Example 8; Page 40; 64pp; English.
CC This sequence represents the tetramerising domain of p53. This sequence,
CC or the oligomerising domains of the yeast transcription factor GCN4 (see
CC W22019 and W22020) can be used in the globin of the invention. The
CC globin of the invention has a non-natural binding domain (BD), preferably
CC an oligomerising domain or ligand binding domain. The globin may be
CC combined with other globins to form a multimeric haemoglobin (Hb). The Hb
CC are used as oxygen carriers, both in vivo (as blood substitutes, volume
CC extenders, in treatment of anaemia and to stimulate haematopoesis) and
CC in vitro (e.g. to improve growth of cell cultures). They are also used to
CC remove oxygen from solutions, or therapeutically to remove nitric oxide.
CC The Hb can also be used as a reference standard for analytical
CC instruments and for delivering drugs or in vivo imaging. Incorporation
CC of the BD allows production of larger Hb that can be assembled without
CC using exogenous crosslinking agents, and the size of the multimer can be
CC controlled. Large Hb show reduced extravasation and prolonged half-life,
CC and are able to deliver oxygen to tissues which erythrocytes can not
CC reach (e.g. downstream of a thrombus, angioplasty balloon etc.).
SQ Sequence 73 AA;

Query Match      100.0%; Score 78; DB 1; Length 73;
Best Local Similarity 100.0%; Pred. No. 6.27e-02;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 27 KPLDGEYFTL 36
QY 1 KPLDGEYFTL 10
|||||

RESULT 3
ID WO9322 standard; peptide; 74 AA.
AC WO9322;
DE 10-JUN-1997 (first entry)
DE C-terminal domain of p53 protein.
KW Chimaeric; bispecific; DNA binding domain; trans; activator; repressor;
KW diphtheria; Pseudomonas; toxin; thymidine kinase; single chain antibody;
KW pathogen; HIV Tat; papilloma virus; B6/E7; Epstein-Barr virus; EBNA;
KW hyperproliferation; p53; tumour; oligomerisation.
OS Homo sapiens.
PN WO9630512-A1.
PD 03-OCT-1996.

Query Match      100.0%; Score 78; DB 1; Length 113;
Best Local Similarity 100.0%; Pred. No. 6.27e-02;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

29-MAR-1996; F00477.
31-MAR-1995; FR-003841.
(RHON ) RHONE POULENC RORER SA.
Bracco L, Schweighoffer F, Tocque B;
WPI; 96-455359/45.
Conditional gene expression system triggered by e.g. infection or
hyper-proliferation - comprises novel bispecific proteins having
DNA-binding domain and second domain specific for trans-activator or
repressor, for gene therapy
Claim 16; Page 44; 81pp; French.
The invention relates to novel chimaeric, bispecific proteins which
comprise: (a) a DNA binding domain and (b) a domain which binds a
trans-activator (TA), trans-repressor (TR) or their complexes, which are
characteristic of a physiological or physiopathological state. The novel
chimaeric, bispecific proteins allow expression of a therapeutic protein
(e.g. diphtheria or Pseudomonas toxins, thymidine kinase, single chain
antibodies) to be regulated in response to particular conditions.
Examples include making the protein responsive to the presence of
particular pathogenic TA mols (e.g. HIV Tat, papilloma virus E6/E7
proteins or Epstein-Barr virus EBNA protein), the therapeutic protein
will be expressed in those cells infected by that pathogen. Similarly,
where the chimaeric protein responds to a cellular protein typical of a
hyperproliferative state (esp. wild-type and mutant p53), expression can
be restricted to tumour cells. The sequence presented here is an example
of a TA binding domain. It corresponds to the C-terminal domain of the
p53 protein between residues 320-393 containing the oligomerisation
domain which binds TA proteins.
SQ Sequence 74 AA;

Query Match      100.0%; Score 78; DB 1; Length 74;
Best Local Similarity 100.0%; Pred. No. 6.27e-02;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

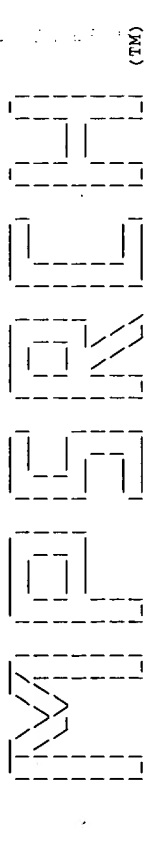
Db 2 KPLDGEYFTL 11
QY 1 KPLDGEYFTL 10
|||||

RESULT 4
ID R51877 standard; Protein; 113 AA.
AC R51877;
DR 18-NOV-1994 (first entry)
DE Human p53 amino acids 237-349.
KW Human nuclear phosphoprotein p53; tumour suppressor gene product;
KW anti-oncogene; cancer; tumour; antibody binding region; epitope.
OS Homo sapiens.
FH Key Location/Qualifiers
FT misc_difference 37 /note= "Arg corresponds to a CAT codon"
FN WO9408241-A.
PD 14-APR-1994.
PF 30-SEP-1993; E02666.
PR 30-SEP-1992; DE-232823
PA (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM STIFTUNG.
PI Klein R, Schranz P, Tessmer C, Volkmann M, Zentgraf H;
DR WPI; 94-135732/16.
DR N-PSDB; Q62362.
PT Non-radioactive detection of p53 specific antibodies - by capture
PT on immobilised p53 or its fragments, then reaction with labelled
PT second antibody, for diagnosis of tumours and suitable for
PT screening
PS Claim 10; Page 19; 35pp; German.
CC Antibodies specific for p53 are detected by binding to immobilised
CC fragments of the p53 gene product containing the antibody-binding
CC region. Preferred fragments contain amino acids 1-241, 40-349,
CC 40-393, 66-241, 66-393, 237-349, 237-393 and esp. 9-33, 37-52 or
CC 368-386. See R51872-R51881 for sequences of these fragments.
SQ Sequence 113 AA;

Query Match      100.0%; Score 78; DB 1; Length 113;
Best Local Similarity 100.0%; Pred. No. 6.27e-02;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

\*\*\*\*\*



(TM)

\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Sat Apr 15 01:04:25 2000; MasPar time 3.13 Seconds  
Tabular output not generated. 75.728 Million cell updates/sec

Title: >US-08-452-843-20  
Description: (1-10) from US08452843.pep  
Perfect Score: 78  
Sequence: 1 KPLDGEYFTL 10

Scoring table: PAM 150  
Gap 15

Searched: 188963 seqs, 23686106 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: a-geneseq36  
1:geneseqp

Statistics: Mean 17.893; Variance 51.195; scale 0.350

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description	Pred. No.
1	78	100.0	48	1 W20035	Human p53 tetramerisat	6.27e-02
2	78	100.0	73	1 W22024	Wild-type p53 tetramer	6.27e-02
3	78	100.0	74	1 W09322	C-terminal domain of p	6.27e-02
4	78	100.0	113	1 R51877	Human p53 amino acids	6.27e-02
5	78	100.0	157	1 R51878	Human p53 amino acids	6.27e-02
6	78	100.0	310	1 R51873	Human p53 amino acids	6.27e-02
7	78	100.0	328	1 R51876	Human p53 amino acids	6.27e-02
8	78	100.0	355	1 W13950	Del356-393 modified hu	6.27e-02
9	78	100.0	363	1 W13975	Modified p53 variant p	6.27e-02
10	78	100.0	363	1 W13972	Modified p53 variant p	6.27e-02
11	78	100.0	363	1 W13954	Modified p53 variant p	6.27e-02
12	78	100.0	363	1 W13976	Modified p53 variant p	6.27e-02
13	78	100.0	363	1 W13971	Modified p53 variant p	6.27e-02
14	78	100.0	374	1 W28482	Human p53 protein vari	6.27e-02
15	78	100.0	374	1 W28481	Human p53 protein vari	6.27e-02
16	78	100.0	381	1 W28480	Human p53 protein vari	6.27e-02
17	78	100.0	381	1 W28489	Human p53 protein vari	6.27e-02
18	78	100.0	393	1 R03191	Amino acid sequence of	6.27e-02
19	78	100.0	393	1 W84270	Human p53 protein	6.27e-02
20	78	100.0	393	1 W69218	Human p53 mutant 1.	6.27e-02
21	78	100.0	393	1 W69217	Human wild-type p53 pr	6.27e-02
22	78	100.0	393	1 W57244	Human p53 protein SEQ	6.27e-02
23	78	100.0	393	1 W57242	Human p53 protein SEQ	6.27e-02

24	78	100.0	393	1 W57243	Human p53 protein SEQ	6.27e-02
25	78	100.0	393	1 W57245	Human p53 protein SEQ	6.27e-02
26	78	100.0	393	1 W13970	Modified p53 variant p	6.27e-02
27	78	100.0	393	1 W13978	Human tumour-derived p	6.27e-02
28	78	100.0	393	1 W13951	Human tumour-derived p	6.27e-02
29	78	100.0	393	1 W13952	Human tumour-derived p	6.27e-02
30	78	100.0	393	1 W02617	Human p53 tumour suppr	6.27e-02
31	78	100.0	393	1 W13979	Human tumour-derived p	6.27e-02
32	78	100.0	393	1 W13948	Human wild-type p53 tu	6.27e-02
33	78	100.0	393	1 R79658	Human p53 protein.	6.27e-02
34	78	100.0	393	1 W05347	Human p53 mutant R248Q	6.27e-02
35	78	100.0	393	1 R91933	Wild type p53 protein.	6.27e-02
36	78	100.0	393	1 W25155	Human p53 variant foun	6.27e-02
37	78	100.0	393	1 W05346	Human p53 mutant R273H	6.27e-02
38	78	100.0	393	1 W05349	Human p53 mutant R273C	6.27e-02
39	78	100.0	393	1 R26758	p53.	6.27e-02
40	78	100.0	393	1 R94623	p53 protein.	6.27e-02
41	78	100.0	401	1 W28487	Human p53 protein vari	6.27e-02
42	78	100.0	401	1 W28488	Human p53 protein vari	6.27e-02
43	78	100.0	438	1 R74272	Tumour suppressor prot	6.27e-02
44	78	100.0	438	1 R50088	p53 tumour suppressor	6.27e-02
45	78	100.0	533	1 W19763	p53-GN-CSF immunostimu	6.27e-02

ALIGNMENTS

RESULT 1  
ID W20035 standard; Protein; 48 AA.  
AC W20035;  
DT 04-SEP-1997 (first entry)  
DE Human p53 tetramerisation domain flanked by linking sequences.  
KW Multimerisation; self assembly; functional domain; linker; folding;  
KW multimerisation domain; post-translational modification; secretion;  
KW interleukin-2; TAFI131; TAFI180; TATA box binding associated factor;  
KW P53; histone; H3; H4; thrombospondin; TSP-4; platelet factor; PFA;  
KW cartilage oligomeric protein; COMP.  
OS Synthetic.  
FH Key  
FT domain  
FT 3. .43  
FT /note= "human p53 tetramerisation domain"  
PN W09637621-A2.  
PD 28-NOV-1996.  
PF 23-MAY-1996; E02230.  
PR 23-MAY-1995; EP-107914.  
PA (MORP-) MORPHOSYS GES PROTEINOPTIMIERUNG MBH.  
PI Hoess A. Pack P;  
DR WPI: 97-021226/02.  
DR N-PSDB; T71287.  
PT Multimerisation devices for self assembly of multifunctional  
PT proteins - used to express recombinant multivalent poly:peptide(s)  
PT by incorporation in a cistron encoding the protein  
PS Claim 3; Fig 3; 64pp; English.  
CC W20035 is the product of an expression cassette encoding the  
CC tetramerisation domain of human p53 (residues 319-360). The cassette was  
CC incorporated into a larger DNA sequence comprising, 5' to 3', a 1st  
CC functional domain; a 1st linker sequence; a multimerisation device;  
CC a 2nd linker sequence; and a 2nd functional domain. The multimerisation  
CC device allows the combination of two or more functional domains in a  
CC structure which is capable of self-multimerisation (at least  
CC trimerisation). Functional domains may, for example, bind to a defined  
CC target, catalyse a reaction, block a receptor binding site, inhibit the  
CC action of another protein or bind to a metal ion. Multimerisation  
CC domains from p53, platelet factor 4, thrombospondin, TSP-4, TATA box  
CC binding associated factors and cartilage oligomeric protein may be  
CC used. The multifunctional proteins can be prepared using standard  
CC recombinant micro-organisms, even though the molecular weight of the  
CC assembled protein exceeds that of the proteins commonly expressed in  
CC bacteria. They have low immunogenicity in humans and carry two or more  
CC functions in a single multimeric structure. Use of a combination of in  
CC vivo expression and in vitro synthesis overcomes prior art problems  
CC due to the differences in folding, secretion and post-translational  
CC modifications for different polypeptides in different hosts.  
SQ Sequence 48 AA;

THIS PAGE BLANK (uspto)

SQ SEQUENCE 393 AA; 43723 MW; DA7D302F CRC32;

Query Match 100.0%; Score 64; DB 4; Length 393;  
Best Local Similarity 100.0%; Pred. No. 6.51e-04;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 315 SPOPKKKPL 323  
|||||  
QY 1 SPOPKKKPL 9

RESULT 13  
ID O08901 PRELIMINARY; PRT; 1058 AA.  
AC O08901;  
DT 01-JUL-1997 (TREMELrel. 04, Created)  
DT 01-JUL-1997 (TREMELrel. 04, Last sequence update)  
DT 01-NOV-1999 (TREMELrel. 12, Last annotation update)  
DE BUDDING INHIBITED BY BENZIMIDAZOLES 1 (S.  
DE CEREBISIAE) HOMOLOG (MITOTIC CHECKPOINT PROTEIN KINASE).  
GN BUB1.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 97325748.  
RA TAYLOR S.S., MCKEON F.;  
RT "Kinetochore localization of murine Bubl1 is required for normal  
RT mitotic timing and checkpoint response to spindle damage.";  
RL Cell 89:727-735(1997).  
DR EMBL; AF002823; AAC53226.1; -.  
DR MGD; MGI:1100510; Bubl1.  
DR PFAM; PF00069; pkinase; 2.  
SQ SEQUENCE 1058 AA; 119562 MW; 73A3AFA5 CRC32;

Query Match 84.4%; Score 54; DB 11; Length 1058;  
Best Local Similarity 87.5%; Pred. No. 1.83e-01;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 528 POPKKKPL 535  
|||||  
QY 2 POPKKKPL 9

RESULT 14  
ID O09007 PRELIMINARY; PRT; 1102 AA.  
AC O09007;  
DT 01-JUL-1997 (TREMELrel. 04, Created)  
DT 01-JUL-1997 (TREMELrel. 04, Last sequence update)  
DT 01-NOV-1999 (TREMELrel. 12, Last annotation update)  
DE BUDDING INHIBITED BY BENZIMIDAZOLES 1 (S.  
DE CEREBISIAE) HOMOLOG (PROTEIN KINASE) (FRAGMENT).  
GN BUB1.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 98110573.  
RA FANGILINAN F., LI Q., WEAVER T., LEWIS B.C., DANG C.V., SPENCER F.;  
RT "Mammalian BUB1 protein kinases: map positions and in vivo  
RT expression.";  
RL Genomics 46:379-388(1997).  
DR EMBL; U89795; AAC53533.1; -.  
DR MGD; MGI:1100510; Bubl1.  
DR PFAM; PF00069; pkinase; 2.  
FT NON\_TER 1  
SQ SEQUENCE 1102 AA; 124320 MW; F69C965F CRC32;

Query Match 84.4%; Score 54; DB 11; Length 1102;  
Best Local Similarity 87.5%; Pred. No. 1.83e-01;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 572 POPKKKPL 579  
|||||  
QY 2 POPKKKPL 9

RESULT 15  
ID O96765 PRELIMINARY; PRT; 256 AA.  
AC O96765;  
DT 01-MAY-1999 (TREMELrel. 10, Created)  
DT 01-MAY-1999 (TREMELrel. 10, Last sequence update)  
DT 01-MAY-1999 (TREMELrel. 10, Last annotation update)  
DE ESAG-9, PUTATIVE PROTEIN (FRAGMENT).  
OS Trypanosoma brucei.  
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TRANSPOSON-INGI;  
RA BARRY J.D.;  
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TRANSPOSON-INGI;  
RX MEDLINE; 91081307.  
RA MATTHEWS K.R., SHIELDS P.G., GRAHAM S.V., COWAN C., BARRY J.D.;  
RT "Duplicative activation mechanisms of two trypanosome telomeric VSG  
RT genes with structurally simple 5' flanks.";  
RL Nucleic Acids Res. 18:7219-7227(1990).  
DR EMBL; AJ012198; CAA09950.1; -.  
FT NON\_TER 1  
SQ SEQUENCE 256 AA; 28105 MW; E3C90BEF3 CRC32;

Query Match 82.8%; Score 53; DB 5; Length 256;  
Best Local Similarity 77.8%; Pred. No. 3.13e-01;  
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 61 PPQPSKKPL 69  
|||||  
QY 1 SPOPKKKPL 9

Search completed: Sat Apr 15 01:01:20 2000  
Job time : 91 secs.



CC REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED  
 CC FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF  
 CC CYCLIN-DEPENDENT KINASES (BY SIMILARITY).

CC -1- SUBCELLULAR LOCATION: NUCLEAR.

DR EMBL; X60019; CAA42634.1; -

DR HSSP; P04637; ISAH.

DR PROSITE; PS00348; P53; 1.

DR PFAM; PF00870; P53; 1.

KW Anti-oncogene; DNA-binding; Transcription regulation; Activator;

FT VARIANT 213 213 Q -> R.

FT NON\_TER 393 393

SEQ SEQUENCE 393 AA; 43684 MW; CB70BD7F CRC32;

Query Match 100.0%; Score 64; DB 4; Length 393;

Best Local Similarity 100.0%; Pred. No. 6.51e-04;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 315 SPQPKKKPL 323

|||||

QY 1 SPQPKKKPL 9

RESULT 10 PRELIMINARY; PRT; 393 AA.

ID Q16811;

AC Q16811;

DT 01-NOV-1996 (TREMELrel. 01, Created)

DT 01-NOV-1996 (TREMELrel. 01, Last sequence update)

DT 01-NOV-1999 (TREMELrel. 12, Last annotation update)

DE CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

OC Eutheria; Primates; Catarrhini; Hominidae; Homo.

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE; 85126934.

RA MATLASHIEWSKI G., LAMB P., PIM D., PEACOCK J., CRAWFORD L.,

RA BENCHIMOL S.;

RT "Isolation and characterization of a human p53 cDNA clone: expression

of the human p53 gene."

RL EMBO J. 3:3257-3262(1984).

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE; 87064416.

RA LAMB P., CRAWFORD L.;

RT "Characterization of the human p53 gene."

RL Mol. Cell. Biol. 6:1379-1385(1986)

CC -1- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT

PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL

CYCLE REGULATION, IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY

REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED

FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF

CYCLIN-DEPENDENT KINASES (BY SIMILARITY).

CC -1- SUBCELLULAR LOCATION: NUCLEAR.

DR EMBL; M13121; AAA59987.1; -

DR EMBL; M13112; AAA59987.1; JOINED.

DR EMBL; M13113; AAA59987.1; JOINED.

DR EMBL; M13114; AAA59987.1; JOINED.

DR EMBL; M13115; AAA59987.1; JOINED.

DR EMBL; M13116; AAA59987.1; JOINED.

DR EMBL; M13117; AAA59987.1; JOINED.

DR EMBL; M13118; AAA59987.1; JOINED.

DR EMBL; M13119; AAA59987.1; JOINED.

DR EMBL; M13120; AAA59987.1; JOINED.

DR HSSP; P04637; ITSR.

DR PROSITE; PS00348; P53; 1.

DR PFAM; PF00870; P53; 1.

KW Repeat; Tumor antigen; Anti-oncogene; DNA-binding;

Transcription regulation; Activator; Nuclear protein; Phosphorylation.

FT NON\_TER 393 393

SEQ SEQUENCE 393 AA; 43698 MW; 3EA71431 CRC32;

Query Match

100.0%; Score 64; DB 4; Length 393;

Best Local Similarity 100.0%; Pred. No. 6.51e-04;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 315 SPQPKKKPL 323

|||||

QY 1 SPQPKKKPL 9

RESULT 11

ID Q15087;

AC Q15087;

DT 01-NOV-1996 (TREMELrel. 01, Created)

DT 01-NOV-1996 (TREMELrel. 01, Last sequence update)

DT 01-NOV-1999 (TREMELrel. 12, Last annotation update)

DE P53 TRANSFORMATION SUPPRESSOR (FRAGMENT).

GN P53.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

OC Eutheria; Primates; Catarrhini; Hominidae; Homo.

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE; 92007731.

RA FARRELL P.J., ALLAN G., SHANAHAN F., VOUSDEN K.H., CROOK T.;

RT "P53 is frequently mutated in Burkitt's lymphoma cell lines."

RL EMBO J. 10:2879-2887(1991).

DR EMBL; X60014; CAA42629.1; -

DR HSSP; P04637; ISAH.

DR PFAM; PF00870; P53; 1.

FT VARIANT 237 237 I -> M.

FT NON\_TER 393 393

SEQ SEQUENCE 393 AA; 43694 MW; 9BB81992 CRC32;

Query Match

100.0%; Score 64; DB 4; Length 393;

Best Local Similarity 100.0%; Pred. No. 6.51e-04;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 315 SPQPKKKPL 323

|||||

QY 1 SPQPKKKPL 9

RESULT 12

ID Q16848;

AC Q16848;

DT 01-NOV-1996 (TREMELrel. 01, Created)

DT 01-NOV-1996 (TREMELrel. 01, Last sequence update)

DT 01-NOV-1999 (TREMELrel. 12, Last annotation update)

DE CELLULAR TUMOR ANTIGEN P53.

GN TP53.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

OC Eutheria; Primates; Catarrhini; Hominidae; Homo.

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE; 87089826.

RA HARRIS N., BRILL E., SHOHAT O., PROKOCIMER M., WOLF D., ARAI N.,

RA ROTTER V.;

RT "Molecular basis for heterogeneity of the human p53 protein."

RL Mol. Cell. Biol. 6:4650-4656(1986).

CC -1- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT

PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL

CYCLE REGULATION, IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY

REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED

FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF

CYCLIN-DEPENDENT KINASES (BY SIMILARITY).

CC -1- SUBCELLULAR LOCATION: NUCLEAR.

DR EMBL; M14694; AAA61211.1; -

DR HSSP; P04637; ITSR.

DR PROSITE; PS00348; P53; 1.

DR PFAM; PF00870; P53; 1.

DR PRINTS; PR00386; P53SUPPRESSR.

KW Nuclear protein; Phosphorylation; Anti-oncogene; DNA-binding;

Transcription regulation; Activator.

```
Db 315 SPQPKKKPL 323
|||||
QY 1 SPQPKKKPL 9

RESULT 6
ID Q16807 PRELIMINARY; PRT; 393 AA.
AC Q16808
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
DE CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).
GN P53.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 92007731.
RA FARRELL P.J., ALLAN G., SHANAHAN F., VOUSDEN K.H., CROOK T.;
RT "p53 is frequently mutated in Burkitt's lymphoma cell lines.";
RL EMBO J. 10:2879-2887(1991).
CC -1- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT
CC PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL
CC CYCLE REGULATION, IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY
CC REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED
CC FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF
CC CYCLIN-DEPENDENT KINASES (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
DR EMBL; X60011; CAA42626.1; -.
DR HSP; P04637; 1SAH.
DR PFAM; PF00870; P53; 1.
KW Anti-oncogene; DNA-binding; Transcription regulation; Activator;
KW Nuclear protein; Phosphorylation.
FT VARIANT 193 193 R -> H.
FT NON_TER 393 393
SQ SEQUENCE 393 AA; 43731 MW; 279BC9CB CRC32;

Query Match 100.0%; Score 64; DB 4; Length 393;
Best Local Similarity 100.0%; Pred. No. 6.51e-04;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 315 SPQPKKKPL 323
|||||
QY 1 SPQPKKKPL 9

RESULT 7
ID Q16808 PRELIMINARY; PRT; 393 AA.
AC Q16809
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
DE CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).
GN P53.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 92007731.
RA FARRELL P.J., ALLAN G., SHANAHAN F., VOUSDEN K.H., CROOK T.;
RT "p53 is frequently mutated in Burkitt's lymphoma cell lines.";
RL EMBO J. 10:2879-2887(1991).
CC -1- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT
CC PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL
CC CYCLE REGULATION, IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY
CC REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED
CC FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF
CC CYCLIN-DEPENDENT KINASES (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
DR EMBL; X60012; CAA42630.1; -.
DR HSP; P04637; 1SAH.
DR PFAM; PF00870; P53; 1.
KW Anti-oncogene; DNA-binding; Transcription regulation; Activator;
KW Nuclear protein; Phosphorylation.
FT VARIANT 193 193 R -> H.
FT NON_TER 393 393
SQ SEQUENCE 393 AA; 43731 MW; 279BC9CB CRC32;

Query Match 100.0%; Score 64; DB 4; Length 393;
Best Local Similarity 100.0%; Pred. No. 6.51e-04;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 315 SPQPKKKPL 323
|||||
QY 1 SPQPKKKPL 9

RESULT 8
ID Q16535 PRELIMINARY; PRT; 393 AA.
AC Q16536
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
DE P53 TRANSFORMATION SUPPRESSOR (FRAGMENT).
GN P53.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 92007731.
RA FARRELL P.J., ALLAN G., SHANAHAN F., VOUSDEN K.H., CROOK T.;
RT "p53 is frequently mutated in Burkitt's lymphoma cell lines.";
RL EMBO J. 10:2879-2887(1991).
DR EMBL; X60017; CAA42632.1; -.
DR EMBL; X60015; CAA42630.1; -.
DR HSP; P04637; 1SAH.
DR PFAM; PF00870; P53; 1.
FT VARIANT 248 248 Q -> R.
FT NON_TER 393 393
SQ SEQUENCE 393 AA; 43684 MW; 239818A9 CRC32;

Query Match 100.0%; Score 64; DB 4; Length 393;
Best Local Similarity 100.0%; Pred. No. 6.51e-04;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 315 SPQPKKKPL 323
|||||
QY 1 SPQPKKKPL 9

RESULT 9
ID Q16809 PRELIMINARY; PRT; 393 AA.
AC Q16809
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
DE CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).
GN P53.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 92007731.
RA FARRELL P.J., ALLAN G., SHANAHAN F., VOUSDEN K.H., CROOK T.;
RT "p53 is frequently mutated in Burkitt's lymphoma cell lines.";
RL EMBO J. 10:2879-2887(1991).
CC -1- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT
CC PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL
CC CYCLE REGULATION, IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY
CC REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED
CC FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF
CC CYCLIN-DEPENDENT KINASES (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
```

O36006;  
AC 01-JAN-1998 (Tremblrel. 05, Created)  
DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)  
DT 01-NOV-1999 (Tremblrel. 12, Last annotation update)  
DE CELLULAR TUMOR ANTIGEN P53.  
GN P53.  
OS Marmota monax (Woodchuck).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Rodentia; Sciurognathi; Sciuridae; Scurinae; Marmota.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 97376996.  
RA FEITELSON M.A., RANGANATHAN P.N., CLAYTON M.M., ZHANG S.M.;  
RT "Partial characterization of the woodchuck tumor suppressor, p53, and  
RT its interaction with woodchuck hepatitis virus X antigen in  
RT hepatocarcinogenesis."  
RL Oncogene 15:327-336(1997).  
CC -1- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT  
CC PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL  
CC CYCLE REGULATION, IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY  
CC REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED  
CC FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF  
CC CYCLIN-DEPENDENT KINASES (BY SIMILARITY).  
CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
DR EMBL; AJ001022; CAA04478.1; -.  
DR HSP; P04637; ITR.  
DR PROSITE; PS00348; P53; 1.  
DR PFAM; PF00870; P53; 1.  
DR PRINTS; PR00386; P53SUPPRESSR.  
KW Anti-oncogene; DNA-binding; Transcription regulation; Activator;  
KW Nuclear protein; Phosphorylation  
SQ SEQUENCE 391 AA; 43468 MW; 95FAB8F2 CRC32;

Query Match 100.0%; Score 64; DB 6; Length 391;  
Best Local Similarity 100.0%; Pred. No. 6.51e-04;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 313 SPQPKKKPL 321  
|||||||  
QY 1 SPQPKKKPL 9

RESULT 3  
ID Q15088 PRELIMINARY; PRT; 393 AA.  
AC Q15088;  
DT 01-NOV-1996 (Tremblrel. 01, Created)  
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)  
DT 01-NOV-1999 (Tremblrel. 12, Last annotation update)  
DE P53 TRANSFORMATION SUPPRESSOR (FRAGMENT).  
GN P53.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 92007731.  
RA FARRELL P.J., ALLAN G., SHANAHAN F., VOUSDEN K.H., CROOK T.;  
RT "p53 is frequently mutated in Burkitt's lymphoma cell lines."  
RL EMO J. 10:2879-2887(1991).  
DR EMBL; X60016; CAA42631.1; -.  
DR HSP; P04637; ISAH.  
DR PFAM; PF00870; P53; 1.  
DR VARIANT 238 238 Y -> C.  
FT NON\_TER 393 393  
SQ SEQUENCE 393 AA; 43713 MW; A01E1523 CRC32;

Query Match 100.0%; Score 64; DB 4; Length 393;  
Best Local Similarity 100.0%; Pred. No. 6.51e-04;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 315 SPQPKKKPL 323  
|||||||  
QY 1 SPQPKKKPL 9

RESULT 4  
ID Q15086 PRELIMINARY; PRT; 393 AA.  
AC Q15086;  
DT 01-NOV-1996 (Tremblrel. 01, Created)  
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)  
DT 01-NOV-1999 (Tremblrel. 12, Last annotation update)  
DE P53 TRANSFORMATION SUPPRESSOR (FRAGMENT).  
GN P53.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 92007731.  
RA FARRELL P.J., ALLAN G., SHANAHAN F., VOUSDEN K.H., CROOK T.;  
RT "p53 is frequently mutated in Burkitt's lymphoma cell lines."  
RL EMO J. 10:2879-2887(1991).  
DR EMBL; X60013; CAA42628.1; -.  
DR HSP; P04637; ISAH.  
DR PFAM; PF00870; P53; 1.  
DR VARIANT 246 246 T -> M.  
FT NON\_TER 393 393  
SQ SEQUENCE 393 AA; 43682 MW; 943B62A3 CRC32;

Query Match 100.0%; Score 64; DB 4; Length 393;  
Best Local Similarity 100.0%; Pred. No. 6.51e-04;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 315 SPQPKKKPL 323  
|||||||  
QY 1 SPQPKKKPL 9

RESULT 5  
ID Q16810 PRELIMINARY; PRT; 393 AA.  
AC Q16810;  
DT 01-NOV-1996 (Tremblrel. 01, Created)  
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)  
DT 01-NOV-1999 (Tremblrel. 12, Last annotation update)  
DE CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).  
GN P53.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 92007731.  
RA FARRELL P.J., ALLAN G., SHANAHAN F., VOUSDEN K.H., CROOK T.;  
RT "p53 is frequently mutated in Burkitt's lymphoma cell lines."  
RL EMO J. 10:2879-2887(1991).  
CC -1- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT  
CC PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL  
CC CYCLE REGULATION, IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY  
CC REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED  
CC FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF  
CC CYCLIN-DEPENDENT KINASES (BY SIMILARITY).  
CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
DR EMBL; X60020; CAA42635.1; -.  
DR HSP; P04637; ISAH.  
DR PROSITE; PS00348; P53; 1.  
DR PFAM; PF00870; P53; 1.  
KW Anti-oncogene; DNA-binding; Transcription regulation; Activator;  
KW Nuclear protein; Phosphorylation.  
FT VARIANT 254 254 D -> N.  
FT VARIANT 254 254 D -> V.  
FT NON\_TER 393 393  
SQ SEQUENCE 393 AA; 43714 MW; 5F914579 CRC32;

Query Match 100.0%; Score 64; DB 4; Length 393;  
Best Local Similarity 100.0%; Pred. No. 6.51e-04;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

\*\*\*\*\*

MPERCH\_PP protein - protein database search, using Smith-Waterman algorithm  
Run on: Sat Apr 15 00:59:49 2000; MasPar time 6.96 Seconds  
Tabular output not generated. 89.608 Million cell updates/sec

Release 3.1A John F. Collins, Biocomputing Research Unit.

Copyright (c) 1993-1998 University of Edinburgh, U.K.

Distribution rights by Oxford Molecular Ltd

\*\*\*\*\*  
Title: >US-08-452-843-19  
Description: (1-9) from US08452843.pep  
Perfect Score: 64  
Sequence: 1 SPQPKKKPL 9  
Scoring table: PAM 150  
Gap 15  
Searched: 225878 seqs, 69334122 residues  
Post-processing: Minimum Watch 0%  
Listing first 45 summaries  
Database: sp\_trembl12  
1:sp\_archaea 2:sp\_bacteria 3:sp\_fungi 4:sp\_human  
5:sp\_invertebrate 6:sp\_mammal 7:sp\_mhc 8:sp\_orcanelle  
9:sp\_phase 10:sp\_plant 11:sp\_rodent 12:sp\_unclassified  
13:sp\_vertebrate 14:sp\_virus  
Statistics: Mean 20.756; Variance 24.932; scale 0.833

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	64	100.0	391	11	Q9WUR6 CELLULAR TUMOR ANTIGEN	6.51e-04
2	64	100.0	391	6	O36006 CELLULAR TUMOR ANTIGEN	6.51e-04
3	64	100.0	393	4	O15088 P53 TRANSFORMATION SUP	6.51e-04
4	64	100.0	393	4	O15086 P53 TRANSFORMATION SUP	6.51e-04
5	64	100.0	393	4	Q16810 CELLULAR TUMOR ANTIGEN	6.51e-04
6	64	100.0	393	4	Q16807 CELLULAR TUMOR ANTIGEN	6.51e-04
7	64	100.0	393	4	Q16808 CELLULAR TUMOR ANTIGEN	6.51e-04
8	64	100.0	393	4	Q16535 P53 TRANSFORMATION SUP	6.51e-04
9	64	100.0	393	4	Q16809 CELLULAR TUMOR ANTIGEN	6.51e-04
10	64	100.0	393	4	Q16811 CELLULAR TUMOR ANTIGEN	6.51e-04
11	64	100.0	393	4	Q15087 P53 TRANSFORMATION SUP	6.51e-04
12	64	100.0	393	4	Q16848 CELLULAR TUMOR ANTIGEN	6.51e-04
13	54	84.4	1058	11	O08901 BUDDING INHIBITED BY B	1.83e-01
14	54	84.4	1102	11	O09007 ESAG-9, PUTATIVE PROTE	3.13e-01
15	53	82.8	256	5	O96765 SIMILAR TO C. ELEGANS	5.33e-01
16	52	81.3	1057	5	O21691 FAP9.2 PROTEIN.	9.04e-01
17	51	79.7	528	10	O22777 BK65A6.1 (FRAGMENT).	1.52e+00
18	50	78.1	754	4	O9Y460 CALCINEURIN BINDING PR	1.52e+00
19	50	78.1	2220	4	O9Y630 CELLULAR TUMOR ANTIGEN	2.55e+00
20	49	76.6	196	6	O29484 CELLULAR TUMOR ANTIGEN	2.55e+00

21	49	76.6	205	11	O35873 CELLULAR TUMOR ANTIGEN	2.55e+00
22	49	76.6	209	2	P73151 HYPOTHETICAL 23.3 KD P	2.55e+00
23	49	76.6	281	6	O29475 CELLULAR TUMOR ANTIGEN	2.55e+00
24	49	76.6	390	11	O70366 CELLULAR TUMOR ANTIGEN	2.55e+00
25	49	76.6	795	5	O91135 RNA BINDING PROTEIN PU	2.55e+00
26	48	75.0	305	3	P78874 FISSION YEAST (FRAGMEN	4.24e+00
27	48	75.0	514	3	O43000 TRANSMEMBRANE TRANSPOR	4.24e+00
28	48	75.0	756	2	O82859 CELLULOSE SYNTHASE SUB	4.24e+00
29	47	73.4	208	14	O9Y2D9 (SUB.SP.NORWALK LIKE V	7.01e+00
30	47	73.4	208	14	O96721 GENOMIC RNA, 3'TERMINA	7.01e+00
31	47	73.4	238	14	O96721 P53 (FRAGMENT).	7.01e+00
32	47	73.4	286	14	P90332 P53 (FRAGMENT).	7.01e+00
33	47	73.4	286	14	P89003 P53 (FRAGMENT).	7.01e+00
34	47	73.4	378	14	P89002 PUTATIVE SERINE/THREON	7.01e+00
35	47	73.4	468	5	O00807 GELATINASE B.	7.01e+00
36	47	73.4	571	13	O9M7L6 REPETITIVE PROLINE-RIC	1.15e+01
37	46	71.9	378	10	O01979 VON WILLEBRAND FACTOR	1.15e+01
38	46	71.9	410	11	O9X0R2 PUTATIVE MITOTIC CHECK	1.15e+01
39	46	71.9	810	4	O43430 MITOTIC CHECKPOINT PRO	1.15e+01
40	46	71.9	1085	4	O43683 MITOTIC CHECKPOINT KIN	1.15e+01
41	46	71.9	1085	4	O40626 PUTATIVE SERINE/THREON	1.15e+01
42	46	71.9	1085	4	O43643 ENAMELIN.	1.15e+01
43	46	71.9	1274	11	O53196 INSULIN-LIKE GROWTH FA	1.88e+01
44	45	70.3	188	13	P81268 FRUCTOSE-6-PHOSPHATE 2	1.88e+01
45	45	70.3	520	11	O35557	

## ALIGNMENTS

RESULT 1  
ID Q9WUR6 PRELIMINARY; PRT; 391 AA.  
AC Q9WUR6;  
DT 01-NOV-1999 (Tremblrel. 12, Created)  
DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)  
DT 01-NOV-1999 (Tremblrel. 12, Last annotation update)  
DE CELLULAR TUMOR ANTIGEN P53.  
GN P53.  
OS Cavia porcellus (Guinea pig).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Rodentia; Hystricognathi; Caviidae; Cavia.  
RN [1]  
RC SEQUENCE FROM N.A.  
RP TISSUE=SPLEEN;  
RX MEDLINE; 99265972.  
RA D'ERCHIA A.M., PESOLE G., TULLO A., SACCONI C., SBISA E.;  
RT "Guinea pig p53 mRNA: identification of new elements in coding and  
RT untranslated regions and their functional and evolutionary  
RT implications."  
RL Genomics 58:50-64(1999).  
CC -!- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT  
CC PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL  
CC CYCLE REGULATION, IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY  
CC REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED  
CC FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF  
CC CYCLIN-DEPENDENT KINASES (BY SIMILARITY).  
CC -!- SUBCELLULAR LOCATION: NUCLEAR.  
DR EMBL; AJ009673; CAB43196.1;  
DR PROSITE; PS00348; P53; 1.  
KW Anti-oncogene; DNA-binding; Transcription regulation; Activator;  
KW Nuclear protein; Phosphorylation.  
SQ SEQUENCE 391 AA; 43288 MW; BFD34AB4 CRC32;

Query Match: 100.0%; Score 64; DB 11; Length 391;

Best Local Similarity 100.0%; Pred. No. 6.51e-04;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 313 SPQPKKKPL 321

QY 1 SPQPKKKPL 9

RESULT 2

ID O36006 PRELIMINARY; PRT; 391 AA.

RL Mol. Cell. Biol. 15:2367-2373(1995).  
 CC -1- SIMILARITY: BELONGS TO THE L7AE FAMILY OF RIBOSOMAL PROTEINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL: X82782; CAA58023.1; -  
 DR FLYBASE: FBgn0014026; Rpl7A.  
 DR PROSITE: PS01082; RIBOSOMAL\_L7AE; 1.  
 DR PFAM: PF01248; Ribosomal\_L7ae; 1.  
 KW Ribosomal protein.  
 SQ SEQUENCE 271 AA; 30677 MW; C804BEFE CRC32;

Query Match 73.4%; Score 47; DB 1; Length 271;  
 Best Local Similarity 75.0%; Pred. No. 2.14e+00;  
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 6 PRPKKPV 13  
 |:|||||  
 QY 2 PQPKKPL 9

Search completed: Sat Apr 15 00:59:32 2000  
 Job time : 43 secs.





Gene 112:247-250(1992).  
[2]  
RN SEQUENCE FROM N.A.  
RA HOU E.W., WISEMAN R.;  
RL Submitted (APR-1994) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES  
CC GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL  
CC CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN  
CC TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION, IT IS A  
CC TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION  
CC BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF  
CC THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.  
CC APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF  
CC BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2  
CC EXPRESSION.  
CC -!- SUBCELLULAR LOCATION: NUCLEAR.  
CC -!- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY  
CC OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED  
CC IN MANY TYPES OF CANCER.  
CC -!- SIMILARITY: BELONGS TO THE P53 FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; M75144; AAA37085.1; -;  
DR EMBL; U07182; AAB41344.1; -;  
DR PIR; JH0633; JH0633.  
DR HSSP; P04637; LYCO.  
DR PROSITE; PS00348; P53; 1.  
DR PFAM; PF00870; P53; 1.  
KW Anti-oncogene; DNA-binding; Transcription regulation; Activator;  
KW Nuclear protein; Phosphorylation; Apoptosis.  
FT DOMAIN 1 77  
FT DOMAIN 78 153  
FT DOMAIN 319 393  
FT DOMAIN 314 326  
FT MOD\_RES 395  
FT CONFLICT 188 188 G -> S (IN REF. 2).  
SQ SEQUENCE 396 AA; 43631 MW; C2668ADE CRC32;  
Query Match 82.88; Score 53; DB 1; Length 396;  
Best Local Similarity 77.8%; Pred. No. 7.49e-02;  
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
Db 318 SPQPKKRL 326  
QY 1 SPQPKKRL 9  
|||||:|  
RESULT 11  
ID P53\_HORSE STANDARD; PRT; 280 AA.  
AC P79892; Q29481;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).  
GN TP53 OR P53.  
OS Equus caballus (Horse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Perissodactyla; Equidae; Equus.  
RN [1]  
RP SEQUENCE OF 1-263 FROM N.A.  
RC TISSUE=SPLEEN;  
RX MEDLINE; 97070350.  
RA PAZZI K.A., KRAEGL S.A., GRIFFEY S.M., THEON A.P., MADEWELL B.R.;  
RT "Analysis of the equine tumor suppressor gene p53 in the normal horse  
RT and in eight cutaneous squamous cell carcinomas.";

Cancer Lett. 107:125-130(1996).  
[2]  
RN SEQUENCE OF 76-280 FROM N.A.  
RX MEDLINE; 96293865.  
RA NASIR I., REID S.W.;  
RT "Nucleotide sequence of exons 5 to 9 of the p53 tumour-suppressor  
RL gene of the horse (Equus caballus).";  
RL DNA Seq. 6:185-187(1996).  
CC -!- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES  
CC GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL  
CC CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN  
CC TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION, IT IS A  
CC TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION  
CC BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF  
CC THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.  
CC APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF  
CC BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2  
CC EXPRESSION (BY SIMILARITY).  
CC -!- SUBCELLULAR LOCATION: NUCLEAR.  
CC -!- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY  
CC OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED  
CC IN MANY TYPES OF CANCER.  
CC -!- SIMILARITY: BELONGS TO THE P53 FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; S83123; AAB46899.1; -;  
DR EMBL; U37120; AAB18936.1; -;  
DR HSSP; P04637; 1SAH.  
DR PROSITE; PS00348; P53; 1.  
DR PFAM; PF00870; P53; 1.  
KW Anti-oncogene; DNA-binding; Transcription regulation; Activator;  
KW Nuclear protein; Phosphorylation; Apoptosis.  
FT NON\_TER 1 1  
FT DOMAIN 262 274  
FT CONFLICT 79 79 T -> A (IN REF. 2).  
FT CONFLICT 83 83 L -> M (IN REF. 2).  
FT CONFLICT 111 111 A -> V (IN REF. 2).  
FT CONFLICT 138 138 G -> A (IN REF. 2).  
FT NON\_TER 280 280  
SQ SEQUENCE 280 AA; 30985 MW; B494F872 CRC32;  
Query Match 76.6%; Score 49; DB 1; Length 280;  
Best Local Similarity 77.8%; Pred. No. 7.19e-01;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
Db 266 SPQPKKPL 274  
QY 1 SPQPKKPL 9  
|||||  
RESULT 12  
ID P53\_CANFA STANDARD; PRT; 381 AA.  
AC Q29537;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 15-DEC-1998 (Rel. 37, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE CELLULAR TUMOR ANTIGEN P53.  
GN TP53 OR P53.  
OS Canis familiaris (Dog).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Carnivora; Fissipedia; Canidae; Canis.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=LEUKOCYTE;  
RX MEDLINE; 96178696.  
RA VELDHOFEN N., MILNER J.;



CC TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION, IT IS A  
CC TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION  
CC BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF  
CC THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.  
CC APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF  
CC BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2  
CC EXPRESSION (BY SIMILARITY).  
CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
CC -1- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY  
CC OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED  
CC IN MANY TYPES OF CANCER.  
CC -1- SIMILARITY: BELONGS TO THE P53 FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC

DR EMBL; X90592; CAA62216.1; -  
DR HSP; P04637; IYCR  
DR PROSITE; PS00348; P53; 1.  
DR PFAM; PF00870; P53; 1.  
KW Anti-oncogene; DNA-binding; Transcription regulation; Activator;  
KW Nuclear protein; Phosphorylation; Apoptosis.  
FT DOMAIN 1 70 ASP/GLU-RICH (ACIDIC).  
FT DOMAIN 308 321 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
FT MOD\_RES 390 390 PHOSPHORYLATION (BY SIMILARITY).  
SQ SEQUENCE 391 AA; 43435 MW; 30A36172 CRC32;

Query Match 87.5%; Score 56; DB 1; Length 391;  
Best Local Similarity 88.9%; Pred. No. 1.28e-02;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 313 SPQTKKKPL 321  
||| |||||  
Qy 1 SPQPKKKPL 9

RESULT 9  
ID P53\_RAT STANDARD; PRT; 391 AA.  
AC P10361; O09168;  
DT 01-MAR-1989 (Rel. 10, Created)  
DT 01-MAR-1989 (Rel. 10, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE CELLULAR TUMOR ANTIGEN P53.  
GN TP53 OR P53.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 89083585.  
RA SOUSSI T.;  
RT "Nucleotide sequence of a cDNA encoding the rat p53 nuclear  
RT oncoprotein.";  
RL Nucleic Acids Res. 16:11384-11384(1988).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 93181268.  
RA HULLA J.E., SCHNEIDER R.P.;  
RT "Structure of the rat p53 tumor suppressor gene.";  
RL Nucleic Acids Res. 21:713-717(1993).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-SPRAGUE-DAWLEY;  
RA MATHUPALA S.P.;  
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES  
CC GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL  
CC CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN

CC TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION, IT IS A  
CC TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION  
CC BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF  
CC THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.  
CC APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF  
CC BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2  
CC EXPRESSION.  
CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
CC -1- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY  
CC OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED  
CC IN MANY TYPES OF CANCER.  
CC -1- SIMILARITY: BELONGS TO THE P53 FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC

DR EMBL; X13058; CAA31457.1; -  
DR EMBL; L07910; AAA41788.1; JOINED.  
DR EMBL; L07904; AAA41788.1; JOINED.  
DR EMBL; L07905; AAA41788.1; JOINED.  
DR EMBL; L07907; AAA41788.1; JOINED.  
DR EMBL; L07908; AAA41788.1; JOINED.  
DR EMBL; L07909; AAA41788.1; JOINED.  
DR EMBL; U90328; AAB80959.1; -  
DR PIR; S02192; S02192.  
DR HSP; P04637; IPET.  
DR PROSITE; PS00348; P53; 1.  
DR PFAM; PF00870; P53; 1.  
KW Anti-oncogene; DNA-binding; Transcription regulation; Activator;  
KW Nuclear protein; Phosphorylation; Apoptosis.  
FT DOMAIN 1 76 ASP/GLU-RICH (ACIDIC).  
FT DOMAIN 77 151 HYDROPHOBIC.  
FT DOMAIN 277 391 HIGHLY BASIC AND MAY BE INVOLVED IN  
FT INTERACTION WITH DNA.  
FT DOMAIN 309 321 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
FT MOD\_RES 390 390 PHOSPHORYLATION (BY SIMILARITY).  
FT VARIANT 103 103 G -> S.  
FT VARIANT 256 256 E -> G.  
FT CONFLICT 174 174 C -> W (IN REF. 2).  
SQ SEQUENCE 391 AA; 43451 MW; E0114C18 CRC32;

Query Match 87.5%; Score 56; DB 1; Length 391;  
Best Local Similarity 88.9%; Pred. No. 1.28e-02;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 313 SPQOKKKPL 321  
||| |||||  
Qy 1 SPQPKKKPL 9

RESULT 10  
ID P53\_MESAU STANDARD; PRT; 396 AA.  
AC Q00366; P97276;  
DT 01-DEC-1992 (Rel. 24, Created)  
DT 01-DEC-1992 (Rel. 24, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE CELLULAR TUMOR ANTIGEN P53.  
GN TP53.  
OS Mesocricetus auratus (Golden hamster).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae; Mesocricetus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-SYRIAN; TISSUE=KIDNEY;  
RX MEDLINE; 92210007.  
RA LEGROS Y., MCINTYRE P., SOUSSI T.;  
RT "The cDNA cloning and immunological characterization of hamster p53.";  
RT

```

RESULT 6
ID P53_MACMU STANDARD; PRT; 393 AA.
AC P56424;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE CELLULAR TUMOR ANTIGEN P53.
GN TP53 OR P53.
OS Macaca mulatta (Rhesus macaque).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Cercopithecoidea; Cercopithecinae;
OC Macaca.
RN [1]
RP SEQUENCE FROM N.A.
RA KHAN M.A., HANSEN C., WELSH J.A., BENNETT W.P.;
RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES
CC GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL
CC CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN
CC TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION, IT IS A
CC TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION
CC BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF
CC THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.
CC APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF
CC BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2
CC EXPRESSION.
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY
CC OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED
CC IN MANY TYPES OF CANCER.
CC -!- SIMILARITY: BELONGS TO THE P53 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U48956; AAB91534.1; -
CC HSSP; P04637; ISAH.
CC PROSITE; PS00348; P53; 1.
CC PFAM; PF00870; P53; 1.
CC KW Anti-oncogene; DNA-binding; Transcription regulation; Activator;
CC Nuclear protein; Phosphorylation; Apoptosis.
CC FT DOMAIN 1 80 ASP/GLU-RICH (ACIDIC).
CC FT MOD_RES 81 150 HYDROPHOBIC.
CC FT DOMAIN 319 393 HIGHLY BASIC AND MAY BE INVOLVED IN
CC FT INTERACTION WITH DNA.
CC FT DOMAIN 311 323 NUCLEAR LOCALIZATION SIGNAL.
CC FT MOD_RES 392 392 PHOSPHORYLATION (BY SIMILARITY).
CC SEQUENCE 393 AA; 43655 MW; 11A9B7F8 CRC32;
Query Match 100.0%; Score 64; DB 1; Length 393;
Best Local Similarity 100.0%; Pred. No. 9.15e-05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 315 SPQPKKKPL 323
QY 1 SPQPKKKPL 9
RESULT 7
ID P53_SHEEP STANDARD; PRT; 382 AA.
AC P51664;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE CELLULAR TUMOR ANTIGEN P53.
GN TP53.
OS Ovis aries (Sheep).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Lagomorpha; Leporidae; Oryctolagus.
RN [1]
RP SEQUENCE FROM N.A.
RA LE GOAS F., MAY P., RONCO P., CARON DE FROMENTEL C.;
RL "cDNA cloning and immunological characterization of rabbit p53.";
RL Gene 185:169-173(1997).
CC -!- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES
CC GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL
CC CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X81705; CAA57349.1; -
CC HSSP; P04637; 1PET.
CC PROSITE; PS00348; P53; 1.
CC PFAM; PF00870; P53; 1.
CC KW Anti-oncogene; DNA-binding; Transcription regulation; Activator;
CC Nuclear protein; Phosphorylation; Apoptosis.
CC FT DOMAIN 1 66 ASP/GLU-RICH (ACIDIC).
CC FT MOD_RES 300 312 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
CC FT MOD_RES 381 381 PHOSPHORYLATION (BY SIMILARITY).
CC SEQUENCE 382 AA; 42809 MW; 0CB99A00 CRC32;
Query Match 87.5%; Score 56; DB 1; Length 382;
Best Local Similarity 88.9%; Pred. No. 1.28e-02;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Db 304 SPQPKKKPL 312
QY 1 SPQPKKKPL 9
RESULT 8
ID P53_RABIT STANDARD; PRT; 391 AA.
AC Q95330;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE CELLULAR TUMOR ANTIGEN P53.
GN TP53.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Lagomorpha; Leporidae; Oryctolagus.
RN [1]
RP SEQUENCE FROM N.A.
RA LE GOAS F., MAY P., RONCO P., CARON DE FROMENTEL C.;
RL "cDNA cloning and immunological characterization of rabbit p53.";
RL Gene 185:169-173(1997).
CC -!- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES
CC GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL
CC CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN

```

Proc. Natl. Acad. Sci. U.S.A. 87:4766-4770(1990).

[9] DEPHOSPHORYLATION BY PP2A.  
RX MEDLINE; 91172186.  
RA SCHEIDTMANN K.H., MUMBY M.C., RUNDELL K., WALTER G.;  
RT "Dephosphorylation of simian virus 40 large-T antigen and p53 protein  
by protein phosphatase 2A: inhibition by small-t antigen.";  
RL Mol. Cell. Biol. 11:1996-2003(1991).

[10] STRUCTURE BY NMR OF 319-360.  
RX MEDLINE; 94294808.  
RA CLORE G.M., OMICHINSKI J.G., SAKAGUCHI K., ZAMBRANO N., SAKAMOTO H.,  
RA APPELLA E., GRONENBORN A.M.; of the oligomerization domain of p53 by  
RT "high-resolution structure of the oligomerization domain of p53 by  
RT multidimensional NMR.";  
RL Science 265:386-391(1994).

[11] STRUCTURE BY NMR OF 325-355.  
RX MEDLINE; 95292092.  
RA LEE W., HARVEY T.S., YIN Y., YAU P., LITCHFIELD D., ARROWSMITH C.H.;  
RT "Solution structure of the tetrameric minimum transforming domain of  
RT p53.";  
RL Nat. Struct. Biol. 1:877-890(1994).

[12] STRUCTURE BY NMR OF 326-354.  
RX MEDLINE; 98026899.  
RA MCCOY M., STAVRIDIS E.S., WATERMAN J.L., WIECZOREK A.M., OPELLA S.J.,  
RA HALAZONETIS T.D.;  
RT "Hydrophobic side-chain size is a determinant of the  
RT three-dimensional structure of the p53 oligomerization domain.";  
RL EMBO J. 16:6230-6236(1997).

[13] X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 94-289.  
RX MEDLINE; 94294806.  
RA CHO Y., GORINA S., JEFFREY P.D., PAVLETICH N.P.;  
RT "Crystal structure of a p53 tumor suppressor-DNA complex:  
RT understanding tumorigenic mutations.";  
RL Science 265:346-353(1994).

[14] X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS) OF 13-29 IN COMPLEX WITH MDM2.  
RX MEDLINE; 97081050.  
RA KUSSIE P.H., GORINA S., MARECHAL V., ELENBAAS B., MOREAU J.,  
RA LEVINE A.J., PAVLETICH N.P.;  
RT "Structure of the MDM2 oncoprotein bound to the p53 tumor suppressor  
RT transactivation domain.";  
RL Science 274:948-953(1996).

[15] X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 97-287 IN COMPLEX WITH 53BP2.  
RX MEDLINE; 97035414.  
RA GORINA S., PAVLETICH N.P.;  
RT "Structure of the p53 tumor suppressor bound to the ankyrin and SH3  
RT domains of 53BP2.";  
RL Science 274:1001-1005(1996).

[16] REVIEW  
RX MEDLINE; 94090335.  
RA HARRIS C.C.;  
RT "p53: at the crossroads of molecular carcinogenesis and risk  
RT assessment.";  
RL Science 262:1980-1981(1993).

[17] REVIEW ON VARIANTS.  
RX MEDLINE; 91289156.  
RA HOOLSTEIN M., SIDRANSKY D., VOGELSTEIN B., HARRIS C.C.;  
RT "p53 mutations in human cancers.";  
RL Science 253:49-53(1991).

[18] REVIEW ON VARIANTS.  
RX MEDLINE; 96271983.  
RA DE VRIES E.M.G., RIQUE D.O., DE VRIES T.N., HARTMANN A., BLASZYK H.,  
RA LIAO D., SOUSSI T., KOVACH J.S., SOMMER S.S.;  
RT "Database of mutations in the p53 and APC tumor suppressor genes  
RT designed to facilitate molecular epidemiological analyses.";

Hum. Mutat. 7:202-213(1996).

[19] VARIANT ARG-72.  
RX MEDLINE; 91153807.  
RA OLSCHWANG S., LAURENT-PUIG P., VASSAL A., SALMON R.-J., THOMAS G.;  
RT "Characterization of a frequent polymorphism in the coding sequence  
RT of the p53 gene in colonic cancer patients and a control  
RT population.";  
RL Hum. Genet. 86:369-370(1991).

[20] VARIANT LFS THR-133.  
RX MEDLINE; 92034774.  
RA LAW J.C., STRONG L.C., CHIDAMBARAM A., FERRELL R.E.;  
RT "A germ line mutation in exon 5 of the p53 gene in an extended cancer  
RT family.";  
RL Cancer Res. 51:6385-6387(1991).

[21] VARIANTS LFS CYS-245; TRP-248; PRO-252 AND LYS-258.  
RX MEDLINE; 91057657.  
RA MALKIN D., LI F.P., STRONG L.C., FRAUMENI J.F. JR., NELSON C.E.,  
RA KIM D.H., KASSEL J., GRYKA M.A., BISCHOFF F.Z., TAINSKY M.A.,  
RA FRIEND S.H.;  
RT "Germ line p53 mutations in a familial syndrome of breast cancer,  
RT sarcomas, and other neoplasms.";  
RL Science 250:1233-1238(1990).

[22] VARIANT LFS ASP-245.  
RX MEDLINE; 91080929.  
RA SRIVASTAVA S., ZOU Z., PIROLLO K., BLATTNER W., CHANG E.H.;  
RT "Germ-line transmission of a mutated p53 gene in a cancer-prone  
RT family with Li-Fraumeni syndrome.";  
RL Nature 348:747-749(1990).

[23] VARIANT LFS LEU-272.  
RX MEDLINE; 92147883.  
RA FELIX C.A., NAU M.M., TAKAHASHI T., MITSUDOMI T., CHIBA I.,  
RA POPLACK D.G., REAMAN G.H., COLE D.E., LETTERIO J.J., WHANG-PENG J.,  
RA KNUDSEN T., MINNA J.D.;  
RT "Hereditary and acquired p53 gene mutations in childhood acute  
RT lymphoblastic leukemia.";  
RL J. Clin. Invest. 89:640-647(1992).

[24] VARIANTS LFS HIS-273 AND VAL-325.  
RX MEDLINE; 92228023.  
RA MALKIN D., JOLLY K.W., BARBIER N., LOOK A.T., FRIEND S.H.,  
RA GEBHARDT M.C., ANDERSEN T.I., BORRESEN A.-L., LI F.P., GARBER J.,  
RA STRONG L.C.;  
RT "Germ-line mutations of the p53 tumor-suppressor gene in children and  
RT young adults with second malignant neoplasms.";  
RL New Engl. J. Med. 326:1309-1315(1992).

[25] VARIANTS BREAST TUMORS GLN-132; SER-249; LYS-280 AND LYS-285.  
RX MEDLINE; 90295284.  
RA BARTEK J., IGGO R., GANNON J., LANE D.P.;  
RT "Genetic and immunochemical analysis of mutant p53 in human breast  
RT cancer cell lines.";  
RL Oncogene 5:893-899(1990).

[26] VARIANTS COLON TUMORS PHE-241 AND HIS-273.  
RX MEDLINE; 91017544.  
RA RODRIGUES N.R., ROWAN A., SMITH M.E.F., KERR I.B., BODMER W.F.,  
RA GANNON J.V., LANE D.P.;

.... remainder of annotations omitted.

Query Match 100.08; Score 64; DB 1; Length 393;

Best Local Similarity 100.08; Pred.No. 9.15e-05;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 315 SPQPKKKPL 323

|||||

QY 1 SPQPKKKPL 9

FT DOMAIN 1 68 ASP/GLU-RICH (ACIDIC).  
FT DOMAIN 81 150 HYDROPHOBIC.  
FT DOMAIN 319 393 HIGHLY BASIC AND MAY BE INVOLVED IN  
INTERACTION WITH DNA.  
FT DOMAIN 311 323 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
FT MOD\_RES 392 392 PHOSPHORYLATION (BY SIMILARITY).  
SQ SEQUENCE 393 AA; 43696 MW; BBE7DC62 CRC32;

Query Match 100.0%; Score 64; DB 1; Length 393;  
Best Local Similarity 100.0%; Pred. No. 9.15e-05;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 315 SPQPKKPL 323  
| | | | |  
Qy 1 SPQPKKPL 9

RESULT 4  
ID P53\_MACFA STANDARD; PRT; 393 AA.  
AC P56423;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 15-JUL-1998 (Rel. 36, Last annotation update)  
DE CELLULAR TUMOR ANTIGEN P53.  
GN TP53 OR P53.  
OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Primates; Catarrhini; Cercopithecoidea; Cercopithecinae;  
OC Macaca.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA KHAN M.A., HANSEN C., WELSH J.A., BENNETT W.P.;  
RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES  
GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL  
CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN  
TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION, IT IS A  
TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION  
BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF  
THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.  
CC APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF  
BAX AND BCL-2 EXPRESSION, OR BY REPRESSION OF BCL-2  
EXPRESSION.  
CC -!- SUBCELLULAR LOCATION: NUCLEAR.  
CC -!- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY  
OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED  
IN MANY TYPES OF CANCER.  
CC -!- SIMILARITY: BELONGS TO THE P53 FAMILY.

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL; U48957; AAB91535.1; .  
DR HSP; P04637; 1SAH.  
DR PROSITE; PS00348; P53; 1.  
DR PFWA; PF00870; P53; 1.  
KW Anti-oncogene; DNA-binding; Transcription regulation; Activator;  
Nuclear protein; Phosphorylation; Apoptosis.  
FT DOMAIN 1 80 ASP/GLU-RICH (ACIDIC).  
FT DOMAIN 81 150 HYDROPHOBIC.  
FT DOMAIN 319 393 HIGHLY BASIC AND MAY BE INVOLVED IN  
INTERACTION WITH DNA.  
FT DOMAIN 311 323 NUCLEAR LOCALIZATION SIGNAL.  
FT MOD\_RES 392 392 PHOSPHORYLATION (BY SIMILARITY).  
SQ SEQUENCE 393 AA; 43678 MW; 2499AC47 CRC32;

Query Match 100.0%; Score 64; DB 1; Length 393;  
Best Local Similarity 100.0%; Pred. No. 9.15e-05;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Db 315 SPQPKKPL 323  
| | | | |  
Qy 1 SPQPKKPL 9  
RESULT 5  
ID P53\_HUMAN STANDARD; PRT; 393 AA.  
AC P04637;  
DT 13-AUG-1987 (Rel. 05, Created)  
DT 01-MAR-1989 (Rel. 10, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE CELLULAR TUMOR ANTIGEN P53 (PHOSPHOPROTEIN P53).  
GN TP53.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA ZAKUT-HOURI R., BIENZ-TADMOR B., GIVOL D., OREN M.;  
RT "Human p53 cellular tumor antigen: cDNA sequence and expression in  
COS cells.";  
RL EMO J. 4:1251-1255(1985).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA LAMB P., CRAWFORD L.;  
RL "Characterization of the human p53 gene.";  
RN [3]  
RP SEQUENCE FROM N.A.  
RA HARLOW E., WILLIAMSON N.M., RALSTON R., HELFMAN D.M., ADAMS T.E.;  
RT "Molecular cloning and in vitro expression of a cDNA clone for human  
cellular tumor antigen p53.";  
RN [4]  
RP TRANSFORMED HYBRIDOMA SV-80 CELL LINE, SEQUENCE FROM N.A.  
RA HARRIS N., BRILL E., SHOHAT O., PROKOCIMER M., WOLF D., ARAI N.,  
ROTER V.;  
RL "Molecular basis for heterogeneity of the human p53 protein.";  
RN [5]  
RP SEQUENCE FROM N.A.  
RA BUCHMAN V.L., CHUMAKOV P.M., NINKINA N.N., SAMARINA O.P.,  
GEORGIEV G.P.;  
RT "A variation in the structure of the protein-coding region of the  
human p53 gene.";  
RN [6]  
RP Gene 70:245-252(1988).  
RN [6]  
RP SEQUENCE OF 101-393 FROM N.A.  
RA MEDLINE; 85126934.  
RA MATLASHENSKI G., LAMB P., PIM D., PEACOCK J., CRAWFORD L.,  
BENCHIMOL S.;  
RT "Isolation and characterization of a human p53 cDNA clone: expression  
of the human p53 gene.";  
RN [7]  
RP EMO J. 3:3257-3262(1984).  
RN [7]  
RP NUCLEAR LOCALIZATION SIGNAL.  
RA MEDLINE; 90191730.  
RA ADDISON C., JENKINS J.R., STURZBECHER H.-W.;  
RT "The p53 nuclear localisation signal is structurally linked to a  
p34cdc2 kinase motif.";  
RN [8]  
RP Oncogene 5:423-426(1990).  
RN [8]  
RP PHOSPHORYLATION BY P60/CDC2 AND CYCLIN B/CDC2.  
RA MEDLINE; 90280455.  
RA BISCHOFF J.R., FRIEDMAN P.N., MARSHAK D.R., PRIVES C., BEACH D.;  
RT "Human p53 is phosphorylated by p60-cdc2 and cyclin B-cdc2.";

DR PFAM; PF00870; P53; 1.  
KW Anti-oncogene; DNA-binding; Transcription regulation; Activator;  
KW Nuclear protein; Phosphorylation.  
FT NON\_TER 1  
FT DOMAIN 289 301 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
FT NON\_TER 314 314  
SQ SEQUENCE 314 AA; 34618 MW; D07F433B CRC32;  
  
Query Match 100.0%; Score 64; DB 1; Length 314;  
Best Local Similarity 100.0%; Pred. No. 9.15e-05;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Db 293 SPQPKKKPL 301  
|||||||  
Qy 1 SPQPKKKPL 9  
  
RESULT 2  
ID P53\_BOVIN STANDARD; PRT; 386 AA.  
AC Q29628;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 15-JUL-1998 (Rel. 36, Last annotation update)  
DE CELLULAR TUMOR ANTIGEN P53.  
GN TP53.  
OS Bos taurus (Bovine), and Bos indicus (Zebu).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC SPICIES-BOVINE; TISSUE=LIVER;  
RX MEDLINE; 95352829.  
RA DEQUIEDT F., KETTMANN R., BURNY A., WILLEMS L.;  
RT "Nucleotide sequence of the bovine P53 tumor-suppressor cDNA.";  
RL DNA Seq. 5:261-264(1995).  
RN [2]  
RP SEQUENCE OF 13-386 FROM N.A.  
RC SPICIES-BOVINE; STRAIN=HOLSTEIN; TISSUE=THYMUS;  
RX MEDLINE; 96401400.  
RA KOMORI H., ISHIGURO N., HORIUCHI M., SHINAGAWA M., AIDA Y.;  
RT "Predominant p53 mutant in enzootic bovine leukemic cell lines.";  
RL Vet. Immunol. Immunopathol. 52:53-63(1996).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC SPECIES-B.INDICUS; STRAIN=BORAN; TISSUE=BLOOD;  
RA BISHOP R.R.P., GOBRIGHT E.E.I.;  
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES  
CC GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL  
CC CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN  
CC TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION, IT IS A  
CC TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION  
CC BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF  
CC THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.  
CC APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF  
CC BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2  
CC EXPRESSION.  
CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
CC -1- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY  
CC OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED  
CC IN MANY TYPES OF CANCER.  
CC -1- SIMILARITY: BELONGS TO THE P53 FAMILY.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL; X81704; CAA57348.1; -

DR EMBL; D49825; BAA08629.1; -  
DR EMBL; U74486; AAB51214.1; -  
DR HSP; P04637; 1YCR.  
DR PROSITE; PS00348; P53; 1.  
DR PFAM; PF00870; P53; 1.  
KW Anti-oncogene; DNA-binding; Transcription regulation; Activator;  
KW Nuclear protein; Phosphorylation; Apoptosis.  
FT DOMAIN 1 59 ASP/GLU-RICH (ACIDIC).  
FT DOMAIN 304 316 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
FT MOD\_RES 385 385 PHOSPHORYLATION (BY SIMILARITY).  
FT CONFLICT 380 380 R -> T (IN REF. 2).  
SQ SEQUENCE 386 AA; 43255 MW; 0322BF3D CRC32;  
  
Query Match 100.0%; Score 64; DB 1; Length 386;  
Best Local Similarity 100.0%; Pred. No. 9.15e-05;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Db 308 SPQPKKKPL 316  
|||||||  
Qy 1 SPQPKKKPL 9  
  
RESULT 3  
ID P53\_CERAE STANDARD; PRT; 393 AA.  
AC P13481;  
DT 01-JAN-1990 (Rel. 13, Created)  
DT 01-JAN-1990 (Rel. 13, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE CELLULAR TUMOR ANTIGEN P53.  
GN TP53.  
OS Cercopithecus aethiops (Green monkey) (Grivet).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Primates; Catarrhini; Cercopithecoidea; Cercopithecinae;  
OC Chlorocebus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=LIVER;  
RX MEDLINE; 90045967.  
RA RIGAUDY P., ECKHARDT W.;  
RT "Nucleotide sequence of a cDNA encoding the monkey cellular  
RT phosphoprotein p53.";  
RL Nucleic Acids Res. 17:8375-8375(1989).  
CC -1- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES  
CC GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL  
CC CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN  
CC TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION, IT IS A  
CC TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION  
CC BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF  
CC THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.  
CC APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF  
CC BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2  
CC EXPRESSION.  
CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
CC -1- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY  
CC OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED  
CC IN MANY TYPES OF CANCER.  
CC -1- SIMILARITY: BELONGS TO THE P53 FAMILY.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL; X15384; CAA34420.1; -  
DR PIR; S06594; S06594.  
DR HSP; P04637; ISAH.  
DR PROSITE; PS00348; P53; 1.  
DR PFAM; PF00870; P53; 1.  
KW Anti-oncogene; DNA-binding; Transcription regulation; Activator;  
KW Nuclear protein; Phosphorylation; Apoptosis.

\*\*\*\*\*  
M O S E R  
(TM)  
\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Sat Apr 15 00:58:49 2000; MasPar time 3.13 Seconds  
Tabular output not generated. 85.836 Million cell updates/sec

Title: >US-08-452-843-19  
Description: (1-9) from US08452843.pep  
Perfect Score: 64  
Sequence: 1 SPQPKKKPL 9

Scoring table: PAM 150  
Gap 15

Searched: 82229 seqs, 29864866 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: swiss-prot38  
1:swissprot

Statistics: Mean 21.738; Variance 23.354; scale 0.931

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES				Pred. No.	
Result No.	Score	Query Match Length	ID Description		
1	64	100.0	1 P53_SPEBE CELLULAR TUMOR ANTIGEN	9.15e-05	
2	64	100.0	1 P53_BOVIN CELLULAR TUMOR ANTIGEN	9.15e-05	
3	64	100.0	1 P53_CERAE CELLULAR TUMOR ANTIGEN	9.15e-05	
4	64	100.0	1 P53_MACFA CELLULAR TUMOR ANTIGEN	9.15e-05	
5	64	100.0	1 P53_HUMAN CELLULAR TUMOR ANTIGEN	9.15e-05	
6	64	100.0	1 P53_MACMU CELLULAR TUMOR ANTIGEN	9.15e-05	
7	56	87.5	1 P53_SHEEP CELLULAR TUMOR ANTIGEN	1.28e-02	
8	56	87.5	1 P53_RABIT CELLULAR TUMOR ANTIGEN	1.28e-02	
9	56	87.5	1 P53_RAT CELLULAR TUMOR ANTIGEN	1.28e-02	
10	53	82.8	1 P53_MESAU CELLULAR TUMOR ANTIGEN	7.49e-02	
11	49	76.6	1 P53_HORSE CELLULAR TUMOR ANTIGEN	7.19e-01	
12	49	76.6	1 P53_CANFA CELLULAR TUMOR ANTIGEN	7.19e-01	
13	49	76.6	1 P53_MOUSE CELLULAR TUMOR ANTIGEN	7.19e-01	
14	49	76.6	1 P53_CRIGR CELLULAR TUMOR ANTIGEN	7.19e-01	
15	47	73.4	1 RL7A_DROME 60S RIBOSOMAL PROTEIN	2.14e+00	
16	47	73.4	1 P53_FELCA CELLULAR TUMOR ANTIGEN	2.14e+00	
17	47	73.4	1 RPL1_TRYBG RETROTRANSPOSABLE ELEM	2.14e+00	
18	46	71.9	1 FL12_SALTY FL12 PROTEIN (FRAGMENT	3.64e+00	
19	46	71.9	1 SPOA_BACSU STAGE 0 SPOULATION PR	3.64e+00	
20	46	71.9	1 YCS0_YEAST HYPOTHETICAL 96.1 KD P	3.64e+00	
21	46	71.9	1 FOR1_MOUSE FORMIN 4 (LIMB DEFORMI	3.64e+00	
22	46	71.9	1 FORM_MOUSE FORMIN (LIMB DEFORMI	3.64e+00	
23	46	71.9	1 CENE_HUMAN CENTROMERIC PROTEIN E	3.64e+00	

ALIGNMENTS				PRT; 314 AA.	
RESULT	ID	P53_SPEBE	STANDARD;		
AC	064662;				
DT	01-NOV-1997	(Rel. 35, Created)			
DT	01-NOV-1997	(Rel. 35, Last sequence update)			
DE	01-NOV-1997	(Rel. 35, Last annotation update)			
DE	CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).				
GN	TP53.				
OS	Spermophilus beecheyi (Beechey ground squirrel).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;				
OC	Eutheria; Rodentia; Sciurognathi; Sciuridae; Scuriurinae; Spermophilus.				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=THYMUS;				
RX	MEDLINE; 95007566.				
RA	RIVKINA M.B.; CULLEN J.M., ROBINSON W.S., MARION P.L.;				
RT	"State of the p53 gene in hepatocellular carcinomas of ground squirrels and woodchucks with past and ongoing infection with hepadnaviruses."				
RL	Cancer Res. 54:5430-5437(1994).				
CC	-!- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION, IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES. APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2 EXPRESSION.				
CC	-!- SUBCELLULAR LOCATION: NUCLEAR.				
CC	-!- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED IN MANY TYPES OF CANCER.				
CC	-!- SIMILARITY: BELONGS TO THE P53 FAMILY.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).				
CC	-----				
DR	EMBL; U43902; AAA85628.1; -				
DR	HSSP; P04837; LYCS.				
DR	PROSITE; PS00348; P53; 1.				

```

#accession      YCR28.
#status         S40970
#molecule_type translation not shown
#residues       203-719, 'T', 721-823, 'V', 825-830, 'A', 832-870 #label CAR
#cross-references EMBL:S47818
REFERENCE
#authors        Cedersberg, H.; Hohmann, S.; Schaaff-Gerstenschlager, I.;
                  Huse, K.; Zimmermann, F.K.
#submission     submitted to the Protein Sequence Database, March 1992
#accession      S19442
#molecule_type DNA
#residues       1-652, 'LSVLI', #label CED
#cross-references EMBL:X59720; MIPS:YCR030c
#note           this was assumed to be the complete sequence of protein
                  YCR030c
                  this sequence has been revised in reference S74291
#accession      S19440
#molecule_type DNA
#residues       731-870 #label CEW
#cross-references EMBL:X59720
#note           this sequence has been revised in reference S74291
                  this was assumed to be protein YCR029c
GENETICS
#map_position   3R
#note           YCR030c
SUMMARY
#length 870 #molecular-weight 96124 #checksum 3413
Query Match      71.9%; Score 46; DB 2; Length 870;
Best Local Similarity 75.0%; Pred.No. 1.06e+01;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
Db 483 PQSKTKPL 490
  ||:||||
QY 2 PQPKKKPL 9

```

Search completed: Sat Apr 15 00:58:31 2000  
 Job time : 17 secs.

##cross-references GB:M10082; NID:g143584; PID:g143585  
##note these authors assume that the codon ATG for Met-29 is the initiator for translation

REFERENCE  
A26068  
#authors Ikeuchi, T.; Kudoh, J.; Tsunasawa, S.  
#journal Mol. Gen. Genet. (1986) 203:371-376  
#title Amino-terminal structure of spo0A protein and sequence homology with spo0F and spo0B proteins.  
#cross-references MUID:86310272  
#contents annotation  
#note initiation at the codon GTG for Met-1 was demonstrated

REFERENCE  
I40013  
#authors Shoji, K.; Hiratsuka, S.; Kawamura, F.; Kobayashi, Y.  
#journal J. Gen. Microbiol. (1988) 134:3249-3257  
#title New Suppressor Mutation sur0B of spo0B and spo0F Mutations in Bacillus subtilis.  
#cross-references MUID:90063528  
#accession I40013  
##status translated from GB/EMBL/DBJ  
##molecule\_type DNA  
##residues 1-16 ##label RES  
##cross-references GB:M23656; NID:g143720; PID:g143721  
A69580  
#authors Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azavedo, V.; Bertero, M.G.; Bessieres, P.; Bolotin, A.; Borchert, S.; Boriss, R.; Boursier, L.; Brans, A.; Braun, C.V.; Brignell, S.C.; Bron, S.; Brouillet, S.; Brusch, M.; Caldwell, B.; Capuano, V.; Carter, N.M.; Choi, S.K.; Codani, J.J.; Connerton, I.F.; Cummings, N.J.; Daniel, R.A.; Denzot, F.; Devine, K.M.; Duesterhoeft, A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.; Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galleron, N.; Ghim, S.Y.; Glaser, P.; Goffeau, A.; Golightly, E.J.; Grandi, G.; Guisepi, G.; Guy, B.J.; Haga, K.; Haeck, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Huilo, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karamata, D.; Kasahara, Y.; Klaerr-Blanchard, M.; Klein, C.; Kobayashi, K.; Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois, S.; Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maveel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno, M.; Moestl, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle, D.; Porwollik, S.; Prescott, A.M.; Presecan, E.; Pujic, P.; Purnelle, B.; Rapoport, G.; Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon, E.; Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror, S.J.; Serror, P.; Shin, B.S.; Soldo, B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Takahashi, H.; Takenaru, K.; Takeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpsira, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.; Vandenbol, M.; Vannier, F.; Vassarotti, A.; Viari, A.; Wambutt, R.; Wedler, E.; Wedler, H.; Weitzengger, T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.; Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.  
#journal Nature (1997) 390:249-256  
#title The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.  
#cross-references MUID:98044033  
#accession A69710  
##status nucleic acid sequence not shown; translation not shown  
##molecule\_type DNA  
##residues 1-267 ##label KUN  
##cross-references GB:Z99116; GB:AL009126; NID:g2634723; PID:e1185690; PID:g2634856  
##experimental-source strain 168  
#COMMENT This protein is involved in the initiation of sporulation.  
GENETICS  
#gene spo0A  
#map\_position 215 (degrees)

##start\_codon GTG  
CLASSIFICATION #superfamily stage 0 sporulation protein A; response regulator homology  
KEYWORDS phosphoprotein; sporulation  
FEATURE  
6-119  
56  
#domain response regulator homology #label RRR\  
#binding\_site phosphate (Asp) (covalent) #status predicted  
SUMMARY #length 267 #molecular-weight 29691 #checksum 9619  
Query Match 71.9%; Score 46; DB 1; Length 267;  
Best Local Similarity 75.0%; Pred. No. 1.06e+01;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
Db 146 PEPKKKL 153  
|:|||||  
QY 2 PPKKKPL 9  
RESULT 14  
ENTRY S14959 #type complete  
TITLE proline-rich protein - wheat  
#formal\_name Triticum aestivum #common\_name common wheat  
DATE 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 09-Sep-1997  
ACCESSIONS S14959  
REFERENCE S14959  
#authors Raines, C.A.; Lloyd, J.C.; Chao, S.; John, U.P.; Murphy, G.J.P.  
#journal Plant Mol. Biol. (1991) 16:663-670  
#title A novel proline-rich protein from wheat.  
#cross-references MUID:91329699  
#accession S14959  
##status preliminary  
##molecule\_type mRNA  
##residues 1-378 ##label RAI  
##cross-references EMBL:X52472; NID:g21841; PID:g21842  
SUMMARY #length 378 #molecular-weight 42119 #checksum 8388  
Query Match 71.9%; Score 46; DB 2; Length 378;  
Best Local Similarity 75.0%; Pred. No. 1.06e+01;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
Db 335 PEPKKPL 342  
|:|||||  
QY 2 PPKKKPL 9  
RESULT 15  
ENTRY S74291 #type complete  
TITLE hypothetical protein YCR030c - yeast (Saccharomycetes cerevisiae)  
ALTERNATE\_NAMES hypothetical protein YCR029c  
ORGANISM #formal\_name Saccharomycetes cerevisiae  
DATE 19-Mar-1997 #sequence\_revision 25-Apr-1997 #text\_change 05-Dec-1997  
ACCESSIONS S74291; S40970; S19442; S19440  
REFERENCE S74288  
#authors Wedler, H.; Wambutt, R.  
#submission submitted to the Protein Sequence Database, September 1996  
#accession S74291  
##molecule\_type DNA  
##residues 1-870 ##label WED  
##cross-references EMBL:X59720; NID:g1907116; PID:e308993; PID:g1907173;  
#note this is a revision to the sequence from reference S19439  
REFERENCE S25336  
#authors Carbone, M.L.A.; Panzeri, L.; Falconi, M.M.; Carcano, C.; Plevani, P.; Lucchini, G.  
#journal Yeast (1992) 8:805-812  
#title Nucleotide sequence of 9.2 kb left of CHY1 on yeast chromosome III from strain AB972: evidence for a Ty insertion and functional analysis of open reading frame



```

389      #binding_site phosphoryl-RNA (Ser) (covalent) #status
        predicted
SUMMARY      #length 390 #molecular-weight 43458 #checksum 1260
Query Match      76.6%; Score 49; DB 1; Length 390;
Best Local Similarity 77.8%; Pred. No. 2.53e+00;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 312 SPQPKKKPL 320
II IIIII
QY 1 SPQPKKKPL 9

RESULT 10
ENTRY JC6176 #type complete
TITLE tumor suppressor protein p53 - Chinese hamster
ORGANISM #formal_name Crictetus griseus #common_name Chinese hamster
DATE 11-Apr-1997 #sequence_revision 09-May-1997 #text_change
08-Sep-1997
ACCESSIONS JC6176
REFERENCE Lee, H.; Lerner, J.M.; Hamlin, J.L.
#authors Gene (1997) 184:177-183
#journal Cloning and characterization of Chinese hamster p53 cDNA.
#title #cross-references MUID:97183659
#contents liver
#accession JC6176
#molecule_type mRNA
#residues 1-393 #label LEE
#cross-references GB:U50395; NID:gl842229; PID:gl842230
COMMENT This protein is a multimer, it plays the central role in a complex
DNA damage-sensing network. It binds to replication factor and
TATA-binding protein, and affects DNA replication, transcription,
and recombination by protein/protein interactions.
GENETICS
#gene p53
CLASSIFICATION #superfamily cellular tumor antigen p53
KEYWORDS liver; tumor
SUMMARY #length 393 #molecular-weight 43362 #checksum 4043
Query Match      76.6%; Score 49; DB 2; Length 393;
Best Local Similarity 77.8%; Pred. No. 2.53e+00;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 315 SPPPKKKTL 323
II IIIII
QY 1 SPQPKKKPL 9

RESULT 11
ENTRY A57416 #type complete
TITLE ribosomal protein L7a, cytosolic - fruit fly (Drosophila
ORGANISM #formal_name Drosophila melanogaster
DATE 08-Feb-1996 #sequence_revision 08-Feb-1996 #text_change
17-Mar-1999
ACCESSIONS A57416
REFERENCE Armes, N.; Fried, M.
#authors Mol. Cell. Biol. (1995) 15:2367-2373
#journal The genomic organization of the region containing the
#title Drosophila melanogaster rpl7a (Surf-3) gene differs from
those of the mammalian and avian surfeit loci.
#cross-references MUID:95257916
#accession A57416
#status preliminary
#molecule_type DNA
#residues 1-272 #label ARM
#cross-references GB:X82782
#note authors translated the codon GCG for residue 112 as Pro,
AAC for residue 116 as Lys, and GTG for residue 117 as
Leu

#gene FlyBase:Rpl7a
#cross-references FlyBase:FBgn0014026
#introns 5/3; 47/1; 170/3
CLASSIFICATION #superfamily rat ribosomal protein L7a
KEYWORDS protein biosynthesis; ribosome
SUMMARY #length 272 #molecular-weight 30732 #checksum 517
Query Match      73.4%; Score 47; DB 2; Length 272;
Best Local Similarity 75.0%; Pred. No. 6.62e+00;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 6 PRPKKKPV 13
I:IIIIII
QY 2 POPKKKPL 9

RESULT 12
ENTRY S14915 #type complete
TITLE hypothetical protein 1 - Trypanosoma brucei gambiense
ORGANISM #formal_name Trypanosoma brucei gambiense
DATE 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change
09-Sep-1997
ACCESSIONS S14915; S08043
REFERENCE Aksoy, S.; Williams, S.; Chang, S.; Richards, F.F.
#authors Nucleic Acids Res. (1990) 18:785-792
#journal SLACS retrotransposon from Trypanosoma brucei gambiense is
#title similar to mammalian LINES.
#cross-references MUID:90192150
#accession S14915
#molecule_type DNA
#residues 1-404 #label AKS
#cross-references EMBL:X17078; NID:gl0533; PID:gl0534
KEYWORDS DNA binding; zinc finger
SUMMARY #length 404 #molecular-weight 45463 #checksum 6169
Query Match      73.4%; Score 47; DB 2; Length 404;
Best Local Similarity 75.0%; Pred. No. 6.62e+00;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 167 TPQPKKKA 174
:IIIIII
QY 1 SPQPKKKP 8

RESULT 13
ENTRY S2B50A #type complete
TITLE stage 0 sporulation protein spo0A - Bacillus subtilis
ALTERNATE_NAMES sporulation initiation two-component response regulator spo0A
ORGANISM #formal_name Bacillus subtilis
DATE 28-Dec-1987 #sequence_revision 28-Dec-1987 #text_change
24-Sep-1998
ACCESSIONS A94036; A22665; I40013; A69710; A26068; A29099; B22665
REFERENCE A94036
#authors Ferrari, F.A.; Trach, K.; LeCoq, D.; Spence, J.; Ferrari, E.;
Hoch, J.A.
#journal Proc. Natl. Acad. Sci. U.S.A. (1985) 82:2647-2651
#title Characterization of the spo0A locus and its deduced product.
#cross-references MUID:85190553
#accession A94036
#molecule_type DNA
#residues 1-267 #label FER
#cross-references GB:M10082; NID:gl43584; PID:gl43585
REFERENCE A22665
#authors Kudoh, J.; Ikeuchi, T.; Kurahashi, K.
#journal Proc. Natl. Acad. Sci. U.S.A. (1985) 82:2665-2668
#title Nucleotide sequences of the sporulation gene spo0A and its
mutant genes of Bacillus subtilis.
#cross-references MUID:85190557
#accession A22665
#molecule_type DNA
#residues 1-267 #label KUD

```

```
##molecule_type mRNA
##residues 1-381 ##label HAN
##cross-references EMBL:M13874; NID:g200202; PID:g200203
##note the nucleotide sequence was submitted to the EMBL Data
Library, July 1988
COMMENT This sequence, produced by alternative splicing of the tenth
intron, lacks the carboxyl-terminal sequence necessary for
covalent attachment of RNA. The function of this minor splice
form is not known.
CLASSIFICATION #superfamily cellular tumor antigen p53
KEYWORDS alternative splicing; phosphoprotein; zinc
FEATURE
1-44 #domain transcription activation #status predicted
#label TRA\
#region conserved region I\
#domain DNA-binding core #status predicted #label DBC\
#region L1 loop\
#region L2 loop\
#region L3 loop\
#region conserved region IV\
#region conserved region V\
#region nuclear location signal\
#region tetramer association\
#binding_site phosphate (Ser) (covalent) #status
predicted\
173,176,235,239 #binding_site zinc (Cys, His, Cys, Cys) #status
predicted\
312 #binding_site phosphate (Ser) (covalent) (by cdc2
kinase) #status predicted
#length 381 #molecular_weight 42498 #checksum 8703
SUMMARY
Query Match 76.6%; Score 49; DB 2; Length 381;
Best Local Similarity 77.98; Pred. No. 2,53e+00;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Db 312 SPQKKKPL 320
II IIIII
QY 1 SPQKKKPL 9
RESULT 9
ENTRY DNMS53 #type complete
TITLE cellular tumor antigen p53 - mouse
ALTERNATE_NAMES oncoprotein p53
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE 28-Aug-1985 #sequence_revision 04-Oct-1996 #text_change
12-Feb-1999
ACCESSIONS A22739; S06336; A02684; S38822; S40014; I48703
REFERENCE A22739
#authors Bienz, B.; Zakut-Houri, R.; Givol, D.; Oren, M.
#journal EMBO J. (1984) 3:2179-2183
#cross-references MUID:85027173
#accession A22739
##molecule_type DNA
##residues 1-134,'V',136-390 ##label BIE
##cross-references GB:X00876; NID:g871420; PID:g871421; GB:X01237;
GB:X01700; NID:g53575; PID:g53576
REFERENCE S06336
#authors Chumakov, P.M.
#journal Bioorg. Khim. (1987) 13:1691-1694
#title Primary structure of DNA complementary to murine oncoprotein
p53 mRNA
#cross-references MUID:88221682
#accession S06336
##status not compared with conceptual translation
##molecule_type mRNA
##residues 1-134,'V',136-390 ##label CHU
REFERENCE A02684
#authors Zakut-Houri, R.; Oren, M.; Bienz, B.; Lavie, V.; Hazum, S.;
Givol, D.
#journal Nature (1983) 306:594-597
```

```
##title A single gene and a pseudogene for the cellular tumour
antigen p53.
#cross-references MUID:84068204
#accession A02684
##molecule_type mRNA
##residues 1-159,'H',161-167,'G',169-233,'I',235-390 ##label ZAK
#cross-references GB:X01237; GB:X01700; NID:g53575
REFERENCE S38822
#authors Arai, N.; Nomura, D.; Yokota, K.; Wolf, D.; Brill, E.;
Shohat, O.; Rotter, V.
#journal Mol. Cell. Biol. (1986) 6:3232-3239
#title Immunologically distinct p53 molecules generated by
alternative splicing.
#cross-references MUID:87064640
#accession S38822
##status preliminary
##molecule_type mRNA
##residues 1-390 ##label ARA1
#cross-references EMBL:M13872; NID:g200198; PID:g200199
#accession S38823
##status preliminary
##molecule_type mRNA
##residues 1-167,'G',169-233,'I',235-390 ##label ARA2
#cross-references EMBL:M13873
REFERENCE S40014
#authors Arai, N.; Nomura, D.; Yokota, K.; Wolf, D.; Brill, E.;
Shohat, O.; Rotter, V.
#submission submitted to the EMBL Data Library, July 1988
#accession S40014
##molecule_type mRNA
##residues 1-167,'G',169-390 ##label ARA3
#cross-references EMBL:M13873; NID:g200200; PID:g200201
REFERENCE I48703
#authors Jenkins, J.R.; Rudge, K.; Redmond, S.; Wade-Evans, A.
#journal Nucleic Acids Res. (1984) 12:5609-5626
#title Cloning and expression analysis of full length mouse cDNA
sequences encoding the transformation associated protein
p53.
#cross-references MUID:84272240
#accession I48703
##status preliminary; translated from GB/EMBL/DDBJ
##molecule_type mRNA
##residues 1-47,'R',49-78,'QW',82-390 ##label RES
#cross-references EMBL:X00741; NID:g53570; PID:g53571
COMMENT This DNA-binding protein plays an essential role in the regulation
of cell division, as it is required for the transition from phase
G0 to G1 of the cell cycle.
COMMENT The tetramer association region may exhibit a beta-turn,
beta-sheet, beta-turn, alpha-helix motif.
CLASSIFICATION #superfamily cellular tumor antigen p53
KEYWORDS apoptosis; cell division control; DNA binding; homotetramer;
phosphoprotein; transcription regulation; tumor suppressor;
zinc
FEATURE
1-44 #domain transcription activation #status predicted
#label TRA\
#region conserved region I\
#domain DNA-binding core #status predicted #label DBC\
#region L1 loop\
#region L2 loop\
#region L3 loop\
#region conserved region IV\
#region conserved region V\
#region nuclear location signal\
#region tetramer association\
#binding_site phosphate (Ser) (covalent) #status
predicted\
#binding_site zinc (Cys, His, Cys, Cys) #status
predicted\
#binding_site phosphate (Ser) (covalent) (by cdc2
kinase) #status predicted\
312
```

```
##cross-references EMBL:X13058; NID:g56828; PID:g56829
REFERENCE S41149
#authors Hulla, J.E.; Schneider, R.P.
#journal Nucleic Acids Res. (1993) 21:713-717
#title Structure of the rat p53 tumor suppressor gene.
#cross-references MUID:93181268
#accession S41149
#status preliminary; nucleic acid sequence not shown;
translation not shown
##molecule_type DNA
##residues 1-173; 'W', 175-391 ##label HUL
##cross-references EMBL:L07909
##note the nucleotide sequence was submitted to the EMBL Data
Library, December 1992

GENETICS
#introns 25/2; 32/3; 123/3; 185/1; 259/2; 305/1; 329/3; 365/2
CLASSIFICATION #superfamily cellular tumor antigen p53
KEYWORDS apoptosis; cell division control; DNA binding; homotetramer;
nucleus; phosphoprotein; transcription regulation; tumor
suppressor; zinc

FEATURE
174,177,236,240 binding_site zinc (Cys, His, Cys, Cys) #status
predicted\
390 binding_site phosphoryl-RNA (Ser) (covalent) #status
predicted\

SUMMARY
#length 391 #molecular-weight 43451 #checksum 7105
Query Match 87.5%; Score 56; DB 2; Length 391;
Best Local Similarity 88.9%; Pred. No. 7,28e-02;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 313 SPQPKKPL 321
||| |||||
Qy 1 SPQPKKPL 9

RESULT 6
ENTRY JH0633
TITLE cellular tumor antigen p53 - golden hamster
ALTERNATE_NAMES tumor-suppressor protein p53
ORGANISM #formal_name Mesocricetus auratus #common_name golden hamster
DATE 17-Aug-1992 #sequence_revision 17-Aug-1992 #text_change
08-Sep-1997
ACCESSIONS JH0633
REFERENCE JH0633
#authors Legros, Y.; McIntyre, P.; Soussi, T.
#journal Gene (1992) 112:247-250
#title The cDNA cloning and immunological characterization of
hamster p53.
#cross-references MUID:92210007
#accession JH0633
##molecule_type mRNA
##residues 1-396 ##label LRG
##cross-references GB:M75144; NID:g191414; PID:g191415
##experimental_source kidney, Strain MP1

GENETICS
#gene p53
CLASSIFICATION #superfamily cellular tumor antigen p53
KEYWORDS apoptosis; cell division control; DNA binding; homotetramer;
nucleus; phosphoprotein; transcription regulation; tumor
suppressor; zinc

FEATURE
179,182,241,245 binding_site zinc (Cys, His, Cys, Cys) #status
predicted\
395 binding_site phosphoryl-RNA (Ser) (covalent) #status
predicted\

SUMMARY
#length 396 #molecular-weight 43631 #checksum 6617
Query Match 82.8%; Score 53; DB 2; Length 396;
Best Local Similarity 77.8%; Pred. No. 3,43e-01;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 318 SPQPKKPL 326
||| |||||
Qy 1 SPQPKKPL 9

##cross-references EMBL:X13058; NID:g56828; PID:g56829
REFERENCE S41149
#authors Hulla, J.E.; Schneider, R.P.
#journal Nucleic Acids Res. (1993) 21:713-717
#title Structure of the rat p53 tumor suppressor gene.
#cross-references MUID:93181268
#accession S41149
#status preliminary; nucleic acid sequence not shown;
translation not shown
##molecule_type DNA
##residues 1-173; 'W', 175-391 ##label HUL
##cross-references EMBL:L07909
##note the nucleotide sequence was submitted to the EMBL Data
Library, December 1992

GENETICS
#introns 25/2; 32/3; 123/3; 185/1; 259/2; 305/1; 329/3; 365/2
CLASSIFICATION #superfamily cellular tumor antigen p53
KEYWORDS apoptosis; cell division control; DNA binding; homotetramer;
nucleus; phosphoprotein; transcription regulation; tumor
suppressor; zinc

FEATURE
174,177,236,240 binding_site zinc (Cys, His, Cys, Cys) #status
predicted\
390 binding_site phosphoryl-RNA (Ser) (covalent) #status
predicted\

SUMMARY
#length 391 #molecular-weight 43451 #checksum 7105
Query Match 87.5%; Score 56; DB 2; Length 391;
Best Local Similarity 88.9%; Pred. No. 7,28e-02;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 313 SPQPKKPL 321
||| |||||
Qy 1 SPQPKKPL 9

RESULT 7
ENTRY S75263
TITLE hypothetical protein sll0983 - Synecocystis sp. (strain PCC
6803)
ORGANISM #formal_name Synecocystis sp.
#variety PCC 6803
DATE 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change
21-Aug-1998
ACCESSIONS S75263
REFERENCE S74322
#authors Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asanizu, E.;
Nakamura, Y.; Miyajima, N.; Hirose, M.; Sugita, M.;
Sasamoto, S.; Kimura, T.; Hosouchi, T.; Matsuno, A.;
Muraki, A.; Nakazaki, N.; Naruo, K.; Okumura, S.; Shimo,
S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.;
Yasuda, M.; Tabata, S.
#journal DNA Res. (1996) 3:109-136
#title Sequence analysis of the genome of the unicellular
cyanobacterium Synecocystis sp. PCC6803. II. Sequence
determination of the entire genome and assignment of
potential protein-coding regions.
#cross-references MUID:97061201
#accession S75263
##status nucleic acid sequence not shown; translation not shown
##molecule_type DNA
##residues 1-209 ##label KAN
##cross-references EMBL:D90904; GB:AB001339; NID:g1652225; PID:d1017910;
PID:g1652254
##note the nucleotide sequence was submitted to the EMBL Data
Library, June 1996

GENETICS
#start_codon GTG
SUMMARY
#length 209 #molecular-weight 23287 #checksum 3304
Query Match 76.6%; Score 49; DB 2; Length 209;
Best Local Similarity 44.4%; Pred. No. 2,53e+00;
Matches 4; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

Db 122 PPHRRRPL 130
||| |||||
Qy 1 SPQPKKPL 9

RESULT 8
ENTRY S38824
TITLE cellular tumor antigen p53, minor splice form - mouse
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change
17-Mar-1999
ACCESSIONS S38824; S35478
REFERENCE S38822
#authors Arai, N.; Nomura, D.; Yokota, K.; Wolf, D.; Brill, E.;
Shohat, O.; Rotter, V.
#journal Mol. Cell. Biol. (1986) 6:3232-3239
#title Immunologically distinct p53 molecules generated by
alternative splicing.
#cross-references MUID:87084640
#accession S38824
##molecule_type mRNA
##residues 1-381 ##label ARA
##cross-references GB:M13874; NID:g200202; PID:g200203
S35478
#authors Han, K.A.; Kulesz-Martin, M.F.
#journal Nucleic Acids Res. (1992) 20:1979-1981
#title Alternatively spliced p53 RNA in transformed and normal cells
of different tissue types.
#cross-references MUID:92253421
#accession S35478
##status nucleic acid sequence not shown; translation not shown
```

```
##accession I38088 translated from GB/EMBL/DDBJ
##status
##molecule_type mRNA
##residues 1-71,'P','73-237','Y',239-393 ##label F07
##cross-references EMBL:X60016; NID:g506444; PID:g506445
##accession I38089 translated from GB/EMBL/DDBJ
##status
##molecule_type mRNA
##residues 1-247,'Q',249-393 ##label F08
##cross-references EMBL:X60017; NID:g506446; PID:g506447
##accession I38090 translated from GB/EMBL/DDBJ
##status
##molecule_type mRNA
##residues 1-71,'P','73-162','H',164-393 ##label F09
##cross-references EMBL:X60018; NID:g506448; PID:g506449
##accession I38091 translated from GB/EMBL/DDBJ
##status
##molecule_type mRNA
##residues 1-212,'Q',214-393 ##label F10
##cross-references EMBL:X60019; NID:g506450; PID:g506451
##accession I38092 translated from GB/EMBL/DDBJ
##status
##molecule_type mRNA
##residues 1-253,'D',255-393 ##label F11
##cross-references EMBL:X60020; NID:g506452; PID:g506453
##note
all sequences submitted to the EMBL/GenBank/DDBJ
databases June 1991

REFERENCE
I38093 Futreal, P.A.; Barrett, J.C.; Wiseman, R.W.
Nucleic Acids Res. (1991) 19:6977
#journal
#title An Alu polymorphism intragenic to the TP53 gene.
#cross-references MUID:92107726
#accession I38093
#status translated from GB/EMBL/DDBJ
##molecule_type DNA
##residues 1-393 ##label FUT
##cross-references EMBL:X34156; NID:g35213; PID:g35214
#accession A44905
#authors Yamada, Y.; Yoshida, T.; Hayashi, K.; Sekiya, T.; Yokota, J.;
Hirohashi, S.; Nakatani, K.; Nakano, H.; Sugimura, T.;
Terada, M.
#journal Cancer Res. (1991) 51:5800-5805
#title p53 gene mutations in gastric cancer metastases and in
gastric cancer cell lines derived from metastases.
#cross-references MUID:92034678
#accession A44905
...
Note: remainder of annotations omitted.

Query Match 100.0%; Score 64; DB 1; Length 393;
Best Local Similarity 100.0%; Pred. No. 9.55e-04;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 315 SPQPKKKPL 323
QY 1 SPQPKKKPL 9
|||||

RESULT 3
ENTRY S06594 #type complete
TITLE cellular tumor antigen p53 - green monkey
ORGANISM #formal_name Cercopithecus aethiops #common_name green
monkey, grivet
DATE 28-Feb-1990 #sequence_revision 28-Feb-1990 #text_change
08-Sep-1997
ACCESSIONS S06594
REFERENCE S06594
#authors Rigaudy, P.; Eckhart, W.
#journal Nucleic Acids Res. (1989) 17:8375
#title Nucleotide sequence of a cDNA encoding the monkey cellular
phosphoprotein p53.
#cross-references MUID:90045967
#accession S06594

##molecule_type mRNA
##residues 1-393 ##label RIG
##cross-references EMBL:X16384; NID:g22795; PID:g22796
CLASSIFICATION superfamily cellular tumor antigen p53
KEYWORDS apoptosis; cell division control; DNA binding; homotetramer;
nucleus; phosphoprotein; transcription regulation; tumor
suppressor; zinc

FEATURE
176,179,238,242 #binding_site zinc (Cys, His, Cys, Cys) #status
predicted\
. 392 #binding_site phosphoryl-RNA (Ser) (covalent) #status
predicted

SUMMARY
#length 393 #molecular-weight 43696 #checksum 4263
Query Match 100.0%; Score 64; DB 2; Length 393;
Best Local Similarity 100.0%; Pred. No. 9.55e-04;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 315 SPQPKKKPL 323
QY 1 SPQPKKKPL 9
|||||

RESULT 4
ENTRY JC6193 #type complete
TITLE tumor suppressor p53 - rabbit
ORGANISM #formal_name Oryctolagus cuniculus #common_name domestic
rabbit
DATE 11-Apr-1997 #sequence_revision 09-May-1997 #text_change
17-Mar-1999
ACCESSIONS JC6193
REFERENCE JC6193
#authors Le Goas, F.; May, P.; Ronco, P.; de Fromental, C.C.
#journal Gene (1997) 185:169-173
#title cDNA cloning and immunological characterization of rabbit
p53.
#cross-references MUID:97208869
#accession JC6193
##molecule_type mRNA
##residues 1-391 ##label LEA
##cross-references EMBL:X50592; NID:g1532043; PID:e194962; PID:g1532044
GENETICS
#gene p53
CLASSIFICATION #superfamily cellular tumor antigen p53
KEYWORDS tumor
SUMMARY #length 391 #molecular-weight 43435 #checksum 4367

Query Match 87.5%; Score 56; DB 2; Length 391;
Best Local Similarity 88.9%; Pred. No. 7.28e-02;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 313 SPQTKKKPL 321
QY 1 SPQPKKKPL 9
|||||

RESULT 5
ENTRY S02192 #type complete
TITLE cellular tumor antigen p53 - rat
ALTERNATE_NAMES gene p53 protein; nuclear oncoprotein p53
ORGANISM #formal_name Rattus norvegicus #common_name Norway rat
DATE 18-Oct-1989 #sequence_revision 18-Oct-1989 #text_change
17-Mar-1999
ACCESSIONS S02192; S41149
REFERENCE S02192
#authors Soussi, T.; de Fromental, C.C.; Breugnot, C.; May, E.
#journal Nucleic Acids Res. (1988) 16:11384
#title Nucleotide sequence of a cDNA encoding the rat p53 nuclear
oncoprotein.
#cross-references MUID:89083585
#accession S02192
##molecule_type mRNA
##residues 1-391 ##label SOU
```

A25397; B25397; S42452; S42453; I38082; I38083; I38084;  
I38085; I38086; I38087; I38088; I38089; I38090; I38091;  
I38092; I38093; A44905; I58354; I78850; I52681; S60153  
A52224  
REFERENCE  
#authors Lamb, P.; Crawford, L.  
#journal Mol. Cell. Biol. (1986) 6:1379-1385  
#title Characterization of the human p53 gene.  
#cross-references MUID:87084416  
#accession A25224  
##molecule\_type DNA  
##residues 1-393 #label LAM  
##cross-references EMBL:X01405; GB:M13121; GB:N00032; NID:g189460;  
PID:g386994

JT0436  
REFERENCE  
#authors Buchman, V.L.; Chumakov, P.M.; Ninkina, N.N.; Samarina, O.P.;  
Georgiev, G.P.  
#journal Gene (1988) 70:245-252  
#title A variation in the structure of the protein-coding region of  
the human p53 gene.  
#cross-references MUID:89108008  
#accession A43073  
##molecule\_type DNA  
##residues 1-393 #label BUC1  
##cross-references EMBL:M22898; NID:g189474  
##note this 72-Arg allele appears to be about 5 times more  
frequent than the 72-Pro allele

JT0436  
REFERENCE  
#authors S4073  
#journal Chumakov, P.M.; Almazov, V.P.; Jenkins, J.R.  
#title Submitted to the EMBL Data Library, August 1990  
#accession S4073  
##molecule\_type DNA  
##residues 1-393 #label CHU  
##cross-references EMBL:X54156; NID:g35213; PID:g35214  
S42669  
REFERENCE  
#authors Matlashewski, G.; Lamb, P.; Pim, D.; Peacock, J.; Crawford,  
L.; Benchimol, S.  
#journal EMBO J. (1984) 3:3257-3262  
#title Isolation and characterization of a human p53 cDNA clone:  
expression of the human p53 gene.  
#cross-references MUID:85126934  
#accession S42669  
##molecule\_type mRNA  
##residues 101-393 #label MKI1  
##cross-references EMBL:X01405; NID:g35215; PID:g642241  
A22837  
REFERENCE  
#authors Zakut-Houri, R.; Blenz-Tadmor, B.; Givol, D.; Oren, M.  
#journal EMBO J. (1985) 4:1251-1255  
#title Human p53 cellular tumor antigen: cDNA sequence and  
expression in COS cells.  
#cross-references MUID:85230577  
#accession A22837  
##molecule\_type mRNA  
##residues 1-71, 'P', 73-393 #label ZAK  
##cross-references EMBL:X02469; EMBL:M60950; NID:g35209; PID:g35210  
A55060  
REFERENCE  
#authors Harlow, E.; Williamson, N.M.; Ralston, R.; Helfman, D.M.;  
Adams, T.E.  
#journal Mol. Cell. Biol. (1985) 5:1601-1610  
#title Molecular cloning and in vitro expression of a cDNA clone for  
human cellular tumor antigen p53.  
#cross-references MUID:85267676  
#accession A55060  
##molecule\_type mRNA  
##residues 1-71, 'P', 73-272, 'H', 274-393 #label HAR  
##cross-references GB:K03199; NID:g189478; PID:g189479  
##experimental\_source clone pr4-2, cell line A431  
REFERENCE  
A93086

Harris, N.; Brill, E.; Shohat, O.; Prokocimer, M.; Wolf, D.;  
Arai, N.; Rotter, V.  
#journal Mol. Cell. Biol. (1986) 6:4650-4656  
#title Molecular basis for heterogeneity of the human p53 protein.  
#cross-references MUID:87089826  
#accession A25397  
##molecule\_type mRNA  
##residues 1-78, 'T', 80-393 #label HAR1  
##cross-references EMBL:M14694; NID:g339813; PID:g339814  
##experimental\_source clone p53-H-1, transformed hybridoma SV-80 cell  
line  
#accession B25397  
##molecule\_type mRNA  
##residues 1-71, 'P', 73-78, 'T', 80-393 #label HAR2  
##cross-references EMBL:M14695; NID:g339815; PID:g339816  
##experimental\_source clone p53-H-19, transformed hybridoma SV-80 cell  
line  
REFERENCE  
S42452  
#authors Matlashewski, G.J.; Tuck, S.; Pim, D.; Lamb, P.; Schneider,  
J.; Crawford, L.V.  
#journal Mol. Cell. Biol. (1987) 7:961-963  
#title Primary structure polymorphism at amino acid residue 72 of  
human p53.  
#cross-references MUID:87144273  
#accession S42452  
##molecule\_type mRNA; DNA  
##residues 66-71, 'P', 73-79 #label MKI2  
##experimental\_source clone lambda C113  
##note 72-Cys was also found, and appears to represent a  
polymorphism  
#accession S42453  
##molecule\_type mRNA; DNA  
##residues 66-79 #label MKI3  
##experimental\_source clone J6K  
REFERENCE  
I38082  
#authors Farrell, P.J.; Allan, G.J.; Shanahan, F.; Vousden, K.H.;  
Crook, T.  
#journal EMBO J. (1991) 10:2879-2887  
#title p53 is frequently mutated in Burkitt's lymphoma cell lines.  
#cross-references MUID:92007731  
#accession I38082  
##status translated from GB/EMBL/DBJ  
##molecule\_type mRNA  
##residues 1-189, 'LLSILSEWKEICVSIWMTETLFDIVWCPMSRLRLALT',  
'VPPSTTTCTVTPAWAA' #label F01  
##cross-references EMBL:X60010; NID:g506432; PID:g506433  
##note deletion of a C nucleotide causes a frameshift at  
position 566  
#accession I38083  
##status translated from GB/EMBL/DBJ  
##molecule\_type mRNA  
##residues 1-192, 'R', 194-393 #label F02  
##cross-references EMBL:X60011; NID:g506434; PID:g506435  
#accession I38084  
##status translated from GB/EMBL/DBJ  
##molecule\_type mRNA  
##residues 1-393 #label F03  
##cross-references EMBL:X60012; NID:g506436; PID:g506437  
#accession I38085  
##status translated from GB/EMBL/DBJ  
##molecule\_type mRNA  
##residues 1-245, 'T', 247-393 #label F04  
##cross-references EMBL:X60013; NID:g506438; PID:g506439  
#accession I38086  
##status translated from GB/EMBL/DBJ  
##molecule\_type mRNA  
##residues 1-236, 'I', 238-393 #label F05  
##cross-references EMBL:X60014; NID:g506440; PID:g506441  
#accession I38087  
##status translated from GB/EMBL/DBJ  
##molecule\_type mRNA  
##residues 1-247, 'Q', 249-393 #label F06  
##cross-references EMBL:X60015; NID:g506442; PID:g506443

\*\*\*\*\*  
W P S R L  
\*\*\*\*\*  
(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Sat Apr 15 00:58:14 2000; MasPar time 3.24 Seconds  
Tabular output not generated. 111.458 Million cell updates/sec

Title: >US-08-452-843-19  
Description: (1-9) from US08452843.pep  
Perfect Score: 64  
Sequence: 1 SPQPKKKPL 9

Scoring table: PAM 150  
Gap 15

Searched: 122810 seqs, 40068593 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: p1r62  
1:p1r1 2:p1r2 3:p1r3 4:p1r4

Statistics: Mean 21.056; Variance 26.395; scale 0.798

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	64	100.0	386	2 S51648	cellular tumor antige	9.55e-04
2	64	100.0	393	1 DNHU53	cellular tumor antige	9.55e-04
3	64	100.0	393	2 S06594	cellular tumor antige	9.55e-04
4	56	87.5	391	2 J06193	tumor suppressor p53	7.28e-02
5	56	87.5	391	2 S02192	cellular tumor antige	7.28e-02
6	53	82.8	396	2 JH0633	cellular tumor antige	3.43e-01
7	49	76.6	209	2 S75263	hypothetical protein	2.53e-00
8	49	76.6	381	2 S38824	cellular tumor antige	2.53e-00
9	49	76.6	390	1 DNMS53	cellular tumor antige	2.53e-00
10	49	76.6	393	2 J06176	tumor suppressor prot	2.53e-00
11	47	73.4	272	2 A57416	ribosomal protein l7a	6.62e-00
12	47	73.4	404	2 S14915	hypothetical protein	6.62e-00
13	46	71.9	267	1 S2BS0A	stage 0 sporulation p	1.06e+01
14	46	71.9	378	2 S14959	proline-rich protein	1.06e+01
15	46	71.9	870	2 S74291	hypothetical protein	1.06e+01
16	46	71.9	1206	2 S24407	formin isoform IV - m	1.06e+01
17	46	71.9	1468	2 S11515	formin - mouse	1.06e+01
18	46	71.9	2663	1 S28261	centromere protein E	1.06e+01
19	45	70.3	155	2 S44012	insulin-like growth f	1.69e+01
20	45	70.3	196	2 S57326	grp-binding protein R	1.69e+01
21	45	70.3	334	2 T02674	hypothetical protein	1.69e+01
22	45	70.3	519	2 J04626	6-phosphofructo-2-kin	1.69e+01
23	45	70.3	532	2 S40983	hypothetical protein	1.69e+01

## ALIGNMENTS

RESULT 1  
ENTRY S51648 #type complete  
TITLE cellular tumor antigen p53 - bovine  
ALTERNATE\_NAMES tumor-suppressor protein p53  
ORGANISM #formal\_name Bos primigenius taurus #common\_name cattle  
DATE 07-May-1995 #sequence\_revision 01-Sep-1995 #text\_change 08-Sep-1997  
ACCESSIONS S51648  
REFERENCE S51648  
#authors Dequiedt, F.; Willems, L.; Burny, A.; Kettmann, R.  
#submission Submitted to the EMBL Data Library, September 1994  
#description Nucleotide sequence of the ovine p53 tumor-suppressor gene  
#accession S51648  
#status preliminary  
#molecule\_type mRNA  
#residues 1-386 #label DEQ  
#cross\_references EMBL:X81704; NID:G602332; PID:G602333  
CLASSIFICATION #superfamily cellular tumor antigen p53  
KEYWORDS apoptosis; cell division control; DNA binding; homotetramer; phosphoprotein; transcription regulation; tumor suppressor; zinc

FEATURE 168,171,231,235 #binding\_site zinc (Cys, His, Cys, Cys) #status predicted  
385 #binding\_site phosphoryl-RNA (Ser) (covalent) #status Predicted  
SUMMARY #length 386 #molecular-weight 43255 #checksum 7025  
Query Match 100.0%; Score 64; DB 2; Length 386;  
Best Local Similarity 100.0%; Pred. No. 9.55e-04;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Db 308 SPQPKKKPL 316  
|||||||  
QY 1 SPQPKKKPL 9  
RESULT 2  
ENTRY DNHU53 #type complete  
TITLE cellular tumor antigen p53 - human  
ALTERNATE\_NAMES cellular phosphoprotein p53; oncoprotein p53; transformation suppressor p53; tumor suppressor p53  
ORGANISM #formal\_name Homo sapiens #common\_name man  
DATE 05-Oct-1988 #sequence\_revision 18-Nov-1994 #text\_change 26-Feb-1999  
ACCESSIONS A25224; A43073; JT0436; S40773; S42669; A22837; A5060;

DR WPI: 97-202618/18.  
PT R284K modified p53 protein having DNA binding ability - useful in  
PT treatment of cancer  
PS Example 1: 59-61: 82pp: English.  
CC Modified p53 variant p53C273del364-393 (W13976) has the tumour-  
CC derived cysteine 273 mutation (see also W13952) and a deletion  
CC of the C-terminal 30 amino acids of wild-type p53 (see also  
CC W13948). Cys273 is a Class I p53 tumour mutation that affects DNA  
CC binding. The C-terminal deletion, introduced by site-directed  
CC mutagenesis of p53 DNA, activates the DNA binding of the p53  
CC tumour mutant. This provides the means for pharmacological rescue  
CC of p53 function in cancer patients. Other modified p53 constructs  
CC (W13949-50, W13953-54, W13968-77) have also been produced. Nucleic  
CC acids coding for modified p53 can be used for cancer gene therapy.  
SQ Sequence 363 AA;

Query Match 100.0%; Score 64; DB 1; Length 363;  
Best Local Similarity 100.0%; Pred. No. 1.05e+00;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 315 SPQPKKKPL 323  
|||||  
QY 1 SPQPKKKPL 9

RESULT 15  
ID W13971 standard; Protein: 363 AA.  
AC W13971;  
DT 25-JUN-1997 (first entry)  
DE Modified p53 variant p53R284del364-393.  
KW p53; tumour suppressor; cancer; therapy; cell proliferation;  
KW apoptosis; protein engineering; DNA binding.  
OS Synthetic.  
PN WO9710843-A1.  
PD 27-MAR-1997.  
PF 20-SEP-1996; U15188.  
PR 22-SEP-1995; US-004802.  
PR 21-AUG-1996; US-697221.  
PA (WIST-) WISTAR INST ANATOMY & BIOLOGY.  
PI Halazonetis TD;  
DR WPI: 97-202618/18.  
PT R284K modified p53 protein having DNA binding ability - useful in  
PT treatment of cancer  
PS Example 1: 51-52: 82pp: English.  
CC Modified p53 variant p53R284del364-393 (W13971) has a Thr284 to Arg  
CC substn. (see also W13949) and a deletion of the C-terminal 30  
CC amino acids. The T284R substitution, introduced by site-directed  
CC mutagenesis of p53 DNA, provides a novel p53-DNA contact between a  
CC phosphate of the DNA backbone and p53. The C-terminal deletion  
CC permits in vitro DNA binding. The variant provides the means for  
CC pharmacological rescue of p53 function in cancer patients. Other  
CC modified p53 constructs (W13949-50, W13953-54, W13968-77) have also  
CC been produced. Nucleic acids coding for modified p53 can be used  
CC for cancer gene therapy.  
SQ Sequence 363 AA;

Query Match 100.0%; Score 64; DB 1; Length 363;  
Best Local Similarity 100.0%; Pred. No. 1.05e+00;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 315 SPQPKKKPL 323  
|||||  
QY 1 SPQPKKKPL 9

Search completed: Sat Apr 15 00:57:57 2000  
Job time : 37 secs.

Db 315 SPQPKKKPL 323  
 |||||  
 QY 1 SPQPKKKPL 9

## RESULT 11

ID W13975 standard; Protein; 363 AA.  
 AC W13975;  
 DT 25-JUN-1997 (first entry)  
 DE Modified p53 variant p53H273R284del364-393.  
 KW p53; tumour suppressor; cancer; therapy; cell proliferation;  
 KW apoptosis; protein engineering; DNA binding.  
 OS Synthetic.  
 PN W09710843-A1.  
 PD 27-MAR-1997.  
 PF 20-SEP-1996; U15188.  
 PR 22-SEP-1995; US-004802.  
 PR 21-AUG-1996; US-697221.  
 PA (WIST-) WISTAR INST ANATOMY & BIOLOGY.  
 PI Halazonetis TD;  
 DR WPI; 97-202618/18.  
 PT R284K modified p53 protein having DNA binding ability - useful in  
 PT treatment of cancer  
 PS Example 1; 58-59; 82pp; English.  
 CC Modified p53 variant p53H273R284del364-393 (W13975) has the tumour-  
 CC derived His273 mutation (see also W13952), a Thr284 to Arg substn.  
 CC (see also W13949) and a deletion of the 30 C-terminal amino acids  
 CC of wild-type p53 (W13948). His273 is a Class I p53 tumour mutation  
 CC that affects DNA binding. The T284R substitution, introduced by  
 CC site-directed mutagenesis of p53 DNA, provides a novel p53-DNA  
 CC contact between a phosphate of the DNA backbone and p53, and  
 CC restores DNA binding. The C-terminal deletion permits in vitro  
 CC DNA binding. The construct provides the means for pharmacological  
 CC rescue of p53 function in cancer patients. Other modified p53  
 CC constructs (W13949-50, W13953-54, W13968-77) have also been  
 CC produced. Nucleic acids coding for modified p53 can be used for  
 CC cancer gene therapy.  
 SQ Sequence 363 AA;

Query Match 100.0%; Score 64; DB 1; Length 363;  
 Best Local Similarity 100.0%; Pred. No. 1.05e+00;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 315 SPQPKKKPL 323  
 |||||  
 QY 1 SPQPKKKPL 9

## RESULT 12

ID W13973 standard; Protein; 363 AA.  
 AC W13973;  
 DT 25-JUN-1997 (first entry)  
 DE Modified p53 variant p53Q248R284del364-393.  
 KW p53; tumour suppressor; cancer; therapy; cell proliferation;  
 KW apoptosis; protein engineering; DNA binding.  
 OS Synthetic.  
 PN W09710843-A1.  
 PD 27-MAR-1997.  
 PF 20-SEP-1996; U15188.  
 PR 22-SEP-1995; US-004802.  
 PR 21-AUG-1996; US-697221.  
 PA (WIST-) WISTAR INST ANATOMY & BIOLOGY.  
 PI Halazonetis TD;  
 DR WPI; 97-202618/18.  
 PT R284K modified p53 protein having DNA binding ability - useful in  
 PT treatment of cancer  
 PS Example 1; 54-56; 82pp; English.  
 CC Modified p53 variant p53Q248R284del364-393 (W13973) has the tumour-  
 CC derived Gln248 mutation (see also W13951), a Thr284 to Arg substn.  
 CC (see also W13949) and a deletion of the 30 C-terminal amino acids  
 CC of wild-type p53 (W13948). Gln248 is a Class I p53 tumour mutation  
 CC that affects DNA binding. The T284R substitution, introduced by  
 CC site-directed mutagenesis of p53 DNA, provides a novel p53-DNA

CC contact between a phosphate of the DNA backbone and p53, and  
 CC restores DNA binding. The C-terminal deletion permits in vitro  
 CC DNA binding. The construct provides the means for pharmacological  
 CC rescue of p53 function in cancer patients. Other modified p53  
 CC constructs (W13949-50, W13953-54, W13968-77) have also been  
 CC produced. Nucleic acids coding for modified p53 can be used for  
 CC cancer gene therapy.  
 SQ Sequence 363 AA;

Query Match 100.0%; Score 64; DB 1; Length 363;  
 Best Local Similarity 100.0%; Pred. No. 1.05e+00;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 315 SPQPKKKPL 323  
 |||||  
 QY 1 SPQPKKKPL 9

## RESULT 13

ID W13974 standard; Protein; 363 AA.  
 AC W13974;  
 DT 25-JUN-1997 (first entry)  
 DE Modified p53 variant p53H273del364-393.  
 KW p53; tumour suppressor; cancer; therapy; cell proliferation;  
 KW apoptosis; protein engineering; DNA binding.  
 OS Synthetic.  
 PN W09710843-A1.  
 PD 27-MAR-1997.  
 PF 20-SEP-1996; U15188.  
 PR 22-SEP-1995; US-004802.  
 PR 21-AUG-1996; US-697221.  
 PA (WIST-) WISTAR INST ANATOMY & BIOLOGY.  
 PI Halazonetis TD;  
 DR WPI; 97-202618/18.  
 PT R284K modified p53 protein having DNA binding ability - useful in  
 PT treatment of cancer  
 PS Example 1; 56-57; 82pp; English.  
 CC Modified p53 variant p53H273del364-393 (W13974) has the tumour-  
 CC derived histidine 273 mutation (see also W13952) and a deletion  
 CC of the C-terminal 30 amino acids of wild-type p53 (see also  
 CC W13948). His273 is a Class I p53 tumour mutation that affects DNA  
 CC binding. The C-terminal deletion, introduced by site-directed  
 CC mutagenesis of p53 DNA, activates the DNA binding of the p53  
 CC tumour mutant. This provides the means for pharmacological rescue  
 CC of p53 function in cancer patients. Other modified p53 constructs  
 CC (W13949-50, W13953-54, W13968-77) have also been produced. Nucleic  
 CC acids coding for modified p53 can be used for cancer gene therapy.  
 SQ Sequence 363 AA;

Query Match 100.0%; Score 64; DB 1; Length 363;  
 Best Local Similarity 100.0%; Pred. No. 1.05e+00;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 315 SPQPKKKPL 323  
 |||||  
 QY 1 SPQPKKKPL 9

## RESULT 14

ID W13976 standard; Protein; 363 AA.  
 AC W13976;  
 DT 25-JUN-1997 (first entry)  
 DE Modified p53 variant p53C273del364-393.  
 KW p53; tumour suppressor; cancer; therapy; cell proliferation;  
 KW apoptosis; protein engineering; DNA binding.  
 OS Synthetic.  
 PN W09710843-A1.  
 PD 27-MAR-1997.  
 PF 20-SEP-1996; U15188.  
 PR 22-SEP-1995; US-004802.  
 PR 21-AUG-1996; US-697221.  
 PA (WIST-) WISTAR INST ANATOMY & BIOLOGY.  
 PI Halazonetis TD;



CC leucine zipper domain at the C-terminal. The p53 variants are  
 CC more active and more stable tumour suppressors and apoptosis-inducing  
 CC agents than wild-type p53 and are active where the wild-type protein  
 CC is not, i.e. they are not inactivated by dominant negative or oncogenic  
 CC mutants, nor by other cellular proteins (because the leucine zipper  
 CC domain prevents formation of inactive mixed oligomers).  
 SQ Sequence 363 AA;

Query Match 100.0%; Score 64; DB 1; Length 363;  
 Best Local Similarity 100.0%; Pred. No. 1.05e+00;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 323 SPQPKKKPL 331  
 |||||  
 QY 1 SPQPKKKPL 9

## RESULT

ID W28480 standard; Protein; 363 AA.  
 AC W28480;  
 DT 25-NOV-1997 (first entry)  
 DE Human p53 protein variant V-325H.  
 KW Leucine zipper domain; L2D; oligomerisation domain; mutant; muten;  
 KW substitution; replacement; transactivation; viral protein VP16; HSV;  
 KW anti-oncogene; hyperproliferation; cancer; restenosis;  
 KW tumour suppression; apoptosis.  
 OS Chimeric - Homo sapiens.  
 OS Chimeric - Herpes simplex virus.  
 OS Synthetic.  
 FH Key Location/Qualifiers  
 FT misc\_difference 189  
 FT /note- "Arg residue at position 182 of wild-type  
 FT p53 has been mutated to His"  
 FT  
 FT  
 PN WO9704092-A1.  
 PD 06-FEB-1997.  
 PF 17-JUL-1996; F01111.  
 PR 19-JUL-1995; FR-008729.  
 PA (RHON ) RHONE FOULENC RORER SA.  
 PI Bracco L, Conseiller E;  
 DR WPI; 97-132633/12.  
 PT New p53 variants e.g. with oligomerisation domain replaced by  
 PT leucine zipper - useful for treating hyper-proliferative disorders,  
 PT esp. cancer and restenosis  
 PS Claim 30; Page -; 133pp; French.  
 CC Claimed variants of protein p53 have at least part of the  
 CC oligomerisation domain deleted and replaced by a leucine zipper  
 CC domain. The mutants preferably also have at least part of the p53  
 CC transactivation domain (amino acids 1-74) deleted and replaced by  
 CC the transactivating domain (TD) from herpes simplex virus viral  
 CC protein VP16 (amino acids 411-490). The present sequence is that of  
 CC a specifically claimed p53 variant designated V-325H and comprising  
 CC the VP16 TD, amino acids 75-325 of human wild-type p53 (but with  
 CC Arg182 replaced by His) and a leucine zipper domain at the C-terminal.  
 CC The p53 variants are more active and more stable tumour suppressors  
 CC and apoptosis-inducing agents than wild-type p53 and are active where  
 CC the wild-type protein is not, i.e. they are not inactivated by dominant  
 CC negative or oncogenic mutants, nor by other cellular proteins (because  
 CC the leucine zipper domain prevents formation of inactive mixed  
 CC oligomers).  
 CC (Note: this sequence does not appear in the specification and has  
 CC been produced by modifying the given sequence of variant V-325).  
 SQ Sequence 363 AA;

Query Match 100.0%; Score 64; DB 1; Length 363;  
 Best Local Similarity 100.0%; Pred. No. 1.05e+00;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 323 SPQPKKKPL 331  
 |||||  
 QY 1 SPQPKKKPL 9

## RESULT

Query Match 100.0%; Score 64; DB 1; Length 363;  
 Best Local Similarity 100.0%; Pred. No. 1.05e+00;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

ID W13954 standard; Protein; 363 AA.  
 AC W13954;  
 DT 25-JUN-1997 (first entry)  
 DE Modified p53 variant (del364-393).  
 KW p53; tumour suppressor; cancer; therapy; cell proliferation;  
 KW apoptosis; protein engineering; DNA binding.  
 OS Synthetic.  
 PN WO9710843-A1.  
 PD 27-MAR-1997.  
 PF 20-SEP-1996; U15188.  
 PR 22-SEP-1995; US-004802.  
 PR 21-AUG-1996; US-697221.  
 PA (WIST-) WISTAR INST ANATOMY & BIOLOGY.  
 PI Halazonetis TD;  
 DR R284K modified p53 protein having DNA binding ability - useful in  
 PT treatment of cancer  
 PS Example 1; 49-51; 82pp; English.  
 CC A modified p53 variant (W13954) comprises wild-type p53 (see  
 CC also W13948) having a deletion of the C-terminal 30 amino acids,  
 CC and is obtd. by site-directed mutagenesis of p53 DNA. Deletion of  
 CC the p53 C-terminal 30 amino acids activates the DNA binding of  
 CC common Class I p53 mutants (see also W13951-52). Novel modified  
 CC p53 variants (W13949-50, W13953-54, W13968-77), some contg.  
 CC C-terminal deletions, provide the means for pharmacological rescue  
 CC of p53 function in cancer patients. Nucleic acids coding for  
 CC modified p53 can be used for cancer gene therapy.  
 SQ Sequence 363 AA;

Query Match 100.0%; Score 64; DB 1; Length 363;  
 Best Local Similarity 100.0%; Pred. No. 1.05e+00;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 315 SPQPKKKPL 323  
 |||||  
 QY 1 SPQPKKKPL 9

## RESULT

ID W13972 standard; Protein; 363 AA.  
 AC W13972;  
 DT 25-JUN-1997 (first entry)  
 DE Modified p53 variant p53Q248del364-393.  
 KW p53; tumour suppressor; cancer; therapy; cell proliferation;  
 KW apoptosis; protein engineering; DNA binding.  
 OS Synthetic.  
 PN WO9710843-A1.  
 PD 27-MAR-1997.  
 PF 20-SEP-1996; U15188.  
 PR 22-SEP-1995; US-004802.  
 PR 21-AUG-1996; US-697221.  
 PA (WIST-) WISTAR INST ANATOMY & BIOLOGY.  
 PI Halazonetis TD;  
 DR WPI; 97-203618/18.  
 PT R284K modified p53 protein having DNA binding ability - useful in  
 PT treatment of cancer  
 PS Example 1; 53-54; 82pp; English.  
 CC Modified p53 variant p53Q248del364-393 (W13972) has the tumour-  
 CC derived glutamine 248 mutation (see also W13951) and a deletion  
 CC of the C-terminal 30 amino acids of wild-type p53 (see also  
 CC W13948). Gln248 is a Class I p53 tumour mutation that affected  
 CC binding. The C-terminal deletion, introduced by site-directed  
 CC mutagenesis of p53 DNA, activates the DNA binding of the p53  
 CC tumour mutant. This provides the means for pharmacological rescue  
 CC of p53 function in cancer patients. Other modified p53 constructs  
 CC (W13949-50, W13953-54, W13968-77) have also been produced. Nucleic  
 CC acids coding for modified p53 can be used for cancer gene therapy.  
 SQ Sequence 363 AA;

Query Match 100.0%; Score 64; DB 1; Length 363;  
 Best Local Similarity 100.0%; Pred. No. 1.05e+00;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CC p53. The chimeric proteins have DNA binding activity and can  
 CC replace lost or insufficient p53 function, providing the means for  
 CC pharmacological rescue of p53 function in cancer patients. Nucleic  
 CC acids coding for modified p53 constructs can be used for cancer  
 CC gene therapy.  
 SQ Sequence 359 AA;

Query Match 100.0%; Score 64; DB 1; Length 359;  
 Best Local Similarity 100.0%; Pred. No. 1.05e+00;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 315 SPQPKKKPL 323  
 QY 1 SPQPKKKPL 9

## RESULT 5

ID W13958 standard; Protein; 361 AA.  
 AC W13958;  
 DT 25-JUN-1997 (first entry)  
 DE Chimeric p53 protein.  
 KW p53; tumour suppressor; cancer; therapy; cell proliferation;  
 KW apoptosis; protein engineering; GCN4; DNA binding.  
 OS Chimeric Homo sapiens;  
 OS Chimeric synthetic.  
 FH Key Location/Qualifiers  
 FT region 1..325  
 FT /label= p53wt  
 FT /note= "amino acids 1-325 of wild-type p53"  
 FT /label= Linker  
 FT region 326..328  
 FT /label= Linker  
 FT region 329..361  
 FT /label= GCN4  
 FT /note= "amino acids 249-281 of GCN4 LZ variant"

PN W09710843-AL.  
 PD 27-MAR-1997.  
 PF 20-SEP-1996; U15188.  
 PR 22-SEP-1995; US-004802.  
 PR 21-AUG-1996; US-697221.  
 PA (WIST-) WISTAR INST ANATOMY & BIOLOGY.  
 PI Halazonetis TD;  
 DR WPI; 97-202618/18.  
 PT R284K modified p53 protein having DNA binding ability - useful in  
 PT treatment of cancer  
 PS Disclosure; Refer to Page 8; 82pp; English.  
 CC Chimeric p53 constructs (W13956-67) comprise N-terminal portions  
 CC of human wild-type p53 tumour suppressor (see also W13948) linked  
 CC to a C-terminal portion of the LZ variant (see also W13955) of  
 CC GCN4 and, in some cases, the C-terminal portion of wild-type  
 CC p53. The chimeric proteins have DNA binding activity and can  
 CC replace lost or insufficient p53 function, providing the means for  
 CC pharmacological rescue of p53 function in cancer patients. Nucleic  
 CC acids coding for modified p53 constructs can be used for cancer  
 CC gene therapy.  
 SQ Sequence 361 AA;

Query Match 100.0%; Score 64; DB 1; Length 361;  
 Best Local Similarity 100.0%; Pred. No. 1.05e+00;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 315 SPQPKKKPL 323  
 QY 1 SPQPKKKPL 9

## RESULT 6

ID W13961 standard; Protein; 361 AA.  
 AC W13961;  
 DT 25-JUN-1997 (first entry)  
 DE Chimeric p53 protein.  
 KW p53; tumour suppressor; cancer; therapy; cell proliferation;  
 KW apoptosis; protein engineering; GCN4; DNA binding.  
 OS Chimeric Homo sapiens;

OS Chimeric synthetic.  
 FH Key Location/Qualifiers  
 FT region 1..323  
 FT /label= p53wt  
 FT /note= "amino acids 1-323 of wild-type p53"  
 FT /label= Linker  
 FT region 324..329  
 FT /label= Linker  
 FT region 330..361  
 FT /label= GCN4  
 FT /note= "amino acids 250-281 of GCN4 LZ variant"

PN W09710843-AL.  
 PD 27-MAR-1997.  
 PF 20-SEP-1996; U15188.  
 PR 22-SEP-1995; US-004802.  
 PR 21-AUG-1996; US-697221.  
 PA (WIST-) WISTAR INST ANATOMY & BIOLOGY.  
 PI Halazonetis TD;  
 DR WPI; 97-202618/18.  
 PT R284K modified p53 protein having DNA binding ability - useful in  
 PT treatment of cancer  
 PS Disclosure; Refer to Page 8; 82pp; English.  
 CC Chimeric p53 constructs (W13956-67) comprise N-terminal portions  
 CC of human wild-type p53 tumour suppressor (see also W13948) linked  
 CC to a C-terminal portion of the LZ variant (see also W13955) of  
 CC GCN4 and, in some cases, the C-terminal portion of wild-type  
 CC p53. The chimeric proteins have DNA binding activity and can  
 CC replace lost or insufficient p53 function, providing the means for  
 CC pharmacological rescue of p53 function in cancer patients. Nucleic  
 CC acids coding for modified p53 constructs can be used for cancer  
 CC gene therapy.  
 SQ Sequence 361 AA;

Query Match 100.0%; Score 64; DB 1; Length 361;  
 Best Local Similarity 100.0%; Pred. No. 1.05e+00;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 315 SPQPKKKPL 323  
 QY 1 SPQPKKKPL 9

## RESULT 7

ID W28479 standard; Protein; 363 AA.  
 AC W28479;  
 DT 25-NOV-1997 (first entry)  
 DE Human p53 protein variant V-325 encoded by PEC114.  
 KW Leucine zipper domain; LZD; oligomerisation domain; mutant; mutein;  
 KW substitution; replacement; transactivation; viral protein VP16; HSV;  
 KW anti-oncogene; hyperproliferation; cancer; restenosis;  
 KW tumour suppression; apoptosis.  
 OS Chimeric - Homo sapiens.  
 OS Chimeric - Herpes simplex virus.  
 OS Synthetic.  
 PN W09704092-AL.  
 PD 06-FEB-1997.  
 PF 17-JUL-1996; F01111.  
 PR 19-JUL-1995; FR-008729.  
 PA (RHON ) RHONE POULENC RORER SA.  
 PI Bracco L; Conseiller E;  
 DR WPI; 97-132633/12.

DR N-PSDB; T86215.  
 DR New p53 variants e.g. with oligomerisation domain replaced by  
 FT leucine zipper - useful for treating hyper-proliferative disorders,  
 FT esp. cancer and restenosis  
 PS Claim 30; Pages 76-78; 133pp; French.  
 CC Claimed variants of protein p53 have at least part of the  
 CC oligomerisation domain deleted and replaced by a leucine zipper  
 CC domain. The mutants preferably also have at least part of the p53  
 CC transactivation domain (amino acids 1-74) deleted and replaced by  
 CC the transactivating domain (TD) from herpes simplex virus viral  
 CC protein VP16 (amino acids 411-490). The present sequence is that of  
 CC a specifically claimed p53 variant designated V-325 and comprising  
 CC the VP16 TD, amino acids 75-325 of human wild-type p53 and a

**Abstract**

\*\*\*\*\*  
MPSRCH\_PP  
(TM)  
\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPSRCH\_PP protein - protein database search, using Smith-Waterman algorithm  
Run on: Sat Apr 15 00:57:20 2000; MasPar time 3.19 Seconds  
Tabular output not generated. 66.807 Million cell updates/sec

Title: >US-08-452-843-19  
Description: (1-9) from US08452843.pep  
Perfect Score: 64  
Sequence: 1 SPQPKKKPL 9

Scoring table: PAM 150  
Gap 15

Searched: 188963 seqs, 23686106 residues

Post-processing: Minimum Match 0\$  
Listing first 45 summaries

Database: a-geneseq36  
1:geneseqp

Statistics: Mean 15.203; Variance 44.993; scale 0.338

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	64	100.0	335	1 W28497	Human p53 protein vari	1.05e+00
2	64	100.0	335	1 W28498	Human p53 protein vari	1.05e+00
3	64	100.0	333	1 W28494	Human p53 protein vari	1.05e+00
4	64	100.0	339	1 W13960	Chimeric p53 protein.	1.05e+00
5	64	100.0	361	1 W13958	Chimeric p53 protein.	1.05e+00
6	64	100.0	361	1 W13961	Chimeric p53 protein.	1.05e+00
7	64	100.0	363	1 W28479	Human p53 protein vari	1.05e+00
8	64	100.0	363	1 W28480	Human p53 protein vari	1.05e+00
9	64	100.0	363	1 W13954	Modified p53 variant (	1.05e+00
10	64	100.0	363	1 W13972	Modified p53 variant p	1.05e+00
11	64	100.0	363	1 W13975	Modified p53 variant p	1.05e+00
12	64	100.0	363	1 W13973	Modified p53 variant p	1.05e+00
13	64	100.0	363	1 W13974	Modified p53 variant p	1.05e+00
14	64	100.0	363	1 W13976	Modified p53 variant p	1.05e+00
15	64	100.0	363	1 W13971	Modified p53 variant p	1.05e+00
16	64	100.0	368	1 W13956	Chimeric p53 protein.	1.05e+00
17	64	100.0	374	1 W28482	Human p53 protein vari	1.05e+00
18	64	100.0	374	1 W28481	Human p53 protein vari	1.05e+00
19	64	100.0	381	1 W28489	Human p53 protein vari	1.05e+00
20	64	100.0	381	1 W28490	Human p53 protein vari	1.05e+00
21	64	100.0	383	1 Y03191	Amino acid sequence of	1.05e+00
22	64	100.0	393	1 W84270	Human p53 protein.	1.05e+00
23	64	100.0	393	1 W69218	Human p53 mutant 1.	1.05e+00

24	64	100.0	393	1 W69217	Human wild-type p53 pr	1.05e+00
25	64	100.0	393	1 W57244	Human p53 protein SEQ	1.05e+00
26	64	100.0	393	1 W05346	Human p53 mutant R273H	1.05e+00
27	64	100.0	393	1 W13968	Modified p53 variant p	1.05e+00
28	64	100.0	393	1 W05347	Human p53 mutant R248Q	1.05e+00
29	64	100.0	393	1 W13969	Modified p53 variant p	1.05e+00
30	64	100.0	393	1 W13970	Modified p53 variant p	1.05e+00
31	64	100.0	393	1 W25155	Human p53 variant foun	1.05e+00
32	64	100.0	393	1 W05349	Human p53 mutant R273C	1.05e+00
33	64	100.0	393	1 R01933	Wild type p53 protein.	1.05e+00
34	64	100.0	393	1 W02617	Human p53 tumour suppr	1.05e+00
35	64	100.0	393	1 W13978	Human tumour-derived p	1.05e+00
36	64	100.0	393	1 W13952	Human tumour-derived p	1.05e+00
37	64	100.0	393	1 W13951	Human tumour-derived p	1.05e+00
38	64	100.0	393	1 W13949	T284R modified human p	1.05e+00
39	64	100.0	401	1 W28487	Human p53 protein vari	1.05e+00
40	64	100.0	401	1 W28488	Human p53 protein vari	1.05e+00
41	64	100.0	404	1 W13963	Chimeric p53 protein.	1.05e+00
42	64	100.0	406	1 W13966	Chimeric p53 protein.	1.05e+00
43	64	100.0	406	1 W13964	Chimeric p53 protein.	1.05e+00
44	64	100.0	411	1 W13967	Chimeric p53 protein.	1.05e+00
45	64	100.0	533	1 W19763	p53-GM-CSF immunostimu	1.05e+00

ALIGNMENTS

RESULT 1  
ID W28497 standard; Protein; 335 AA.  
AC W28497;  
DT 25-NOV-1997 (first entry)  
DE Human p53 protein variant 360h-325 encoded by p53179.  
KW Leucine zipper domain; LZD; oligomerisation domain; mutant; mutuin;  
KW substitution; replacement; transactivation; hinge region;  
KW anti-oncogene; hyperproliferation; cancer; restenosis;  
KW tumour suppression; apoptosis.  
OS Homo sapiens.  
FH Synthetic.  
FH Key Location/Qualifiers  
FT region 39..53  
FT /label= hinge  
PN W09704092-A1.  
PD 06-FEB-1997.  
PF 17-JUL-1996; F01111.  
PR 19-JUL-1995; FR-008729.  
PA (RHON ) RHONE POULENC RORER SA.  
PI Bracco L, Conseiller E;  
DR WPI: 97-132633/12.  
DR N-PSDB; T86224.  
PT New p53 variants e.g. with oligomerisation domain replaced by  
PT leucine zipper - useful for treating hyper-proliferative disorders,  
PT esp. cancer and restenosis  
PS Claim 39: Pages 94-95; 133pp; French.  
CC Claimed variants of protein p53 have at least part of the  
CC oligomerisation domain deleted and replaced by a leucine zipper  
CC domain. The mutants preferably also have at least part of the p53  
CC transactivation domain (amino acids 1-74) deleted and replaced by  
CC the domain 325-360 of p53. The present sequence is that of a  
CC specifically claimed p53 variant designated 360h-325 and comprising  
CC the 325-360 domain, separated from amino acids 75-325 of human  
CC wild-type p53 by a synthetic hinge sequence (Gly4Ser)3, and with a  
CC leucine zipper domain at the C-terminal. The p53 variants are  
CC more active and more stable tumour suppressors and apoptosis-inducing  
CC agents than wild-type p53 and are active where the wild-type protein  
CC is not, i.e. they are not inactivated by dominant negative or oncogenic  
CC mutants, nor by other cellular proteins (because the leucine zipper  
CC domain prevents formation of inactive mixed oligomers).  
SQ Sequence 335 AA;

Query Match 100.0%; Score 64; DB 1; Length 335;  
Best Local Similarity 100.0%; Pred. No. 1.05e+00;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
DB 295 SPQPKKKPL 303

**THIS PAGE BLANK (uspto)**

SQ SEQUENCE 281 AA; 31762 MW; FC7BAE31 CRC32;  
Query Match 88.2%; Score 60; DB 6; Length 281;  
Best Local Similarity 100.0%; Pred. No. 4.52e-03;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 193 PPGSTKRAL 201  
QY 2 PPGSTKRAL 10  
|||||

RESULT 13  
ID Q99659; PRELIMINARY; PRT; 45 AA.  
AC Q99659;  
DT 01-MAY-1997 (TREMELrel. 03, Created)  
DT 01-MAY-1997 (TREMELrel. 03, Last sequence update)  
DT 01-NOV-1999 (TREMELrel. 12, Last annotation update)  
DE CELLULAR PHOSPHOPROTEIN P53 (FRAGMENT).  
GN P53.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA FILIPPINI G., SOLDATI G.;  
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.  
DR EMBL; U63714; AAB39322.1; -.  
DR HSSP; P04637; 1TSR.  
DR PFAM; PF00870; P53; 1.  
FT NON\_TER 1 1  
FT NON\_TER 45 45  
SQ SEQUENCE 45 AA; 5170 MW; 09281164 CRC32;

Query Match 82.4%; Score 56; DB 4; Length 45;  
Best Local Similarity 100.0%; Pred. No. 5.29e-02;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 38 LPPGSTKR 45  
QY 1 LPPGSTKR 8  
|||||

RESULT 14  
ID O00392; PRELIMINARY; PRT; 648 AA.  
AC O00392;  
DT 01-JUL-1997 (TREMELrel. 04, Created)  
DT 01-JUL-1997 (TREMELrel. 04, Last sequence update)  
DT 01-NOV-1999 (TREMELrel. 12, Last annotation update)  
DE UV RADIATION RESISTANCE ASSOCIATED PROTEIN.  
GN UVRAG OR UVRAGL.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA MEDLINE; 97312697.  
RX PERELMAN B., DAFNI N., NAIMAN T., ELI D., YAAKOV M., FENG T.L.Y.,  
RA SINHA S., WEBER G., KHODAEI S., SANCAR A., DOTAN I., CANAANI D.;  
RT "Molecular cloning of a novel human gene encoding a 63-kDa protein and  
its sublocalization within the 11q13 locus."  
RL Genomics 41:397-405(1997).  
RN [2]  
RP REVISIONS.  
RA CANAANI D.;  
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL; X99050; CAA67507.1; -.  
DR MIM; 602493; -.  
DR PFAM; PF00168; C2; 1.  
SQ SEQUENCE 648 AA; 72363 MW; 7877028C CRC32;

Query Match 80.9%; Score 55; DB 4; Length 648;  
Best Local Similarity 80.0%; Pred. No. 9.63e-02;  
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 22 LPPGSAARAL 31  
QY 1 LPPGSTKRAL 10  
|||||

RESULT 15  
ID O35873; PRELIMINARY; PRT; 205 AA.  
AC O35873;  
DT 01-JAN-1998 (TREMELrel. 05, Created)  
DT 01-JAN-1998 (TREMELrel. 05, Last sequence update)  
DT 01-NOV-1999 (TREMELrel. 12, Last annotation update)  
DE CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).  
GN P53.  
OS Cricetus griseus (Chinese hamster).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae; Cricetulus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA RAINALDI G., MARCHETTI S., CAPECCHI B., MENEVERI R., PIRAS A.,  
RA LEUZZI R.;  
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA VATERONI L., MUSIO A., MENEVERI R., RAINALDI G.;  
RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT  
CC PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL  
CC CYCLE REGULATION. IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY  
CC REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED  
CC FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF  
CC CYCLIN-DEPENDENT KINASES (BY SIMILARITY).  
CC -1- SUBCELLULAR LOCATION: NUCLEAR.

DR EMBL; U74487; AAB82420.1; -.  
DR HSSP; P04637; 1SAH.  
DR PROSITE; PS00348; P53; 1.  
DR PFAM; PF00870; P53; 1.  
KW Anti-oncogene; DNA-binding; Transcription regulation; Activator;  
KW Nuclear protein; Phosphorylation.  
FT NON\_TER 1 1  
FT NON\_TER 205 205  
SQ SEQUENCE 205 AA; 23122 MW; 680DDDDC CRC32;

Query Match 79.4%; Score 54; DB 11; Length 205;  
Best Local Similarity 80.0%; Pred. No. 1.74e-01;  
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 174 LPPGSAARAL 183  
QY 1 LPPGSTKRAL 10  
|||||

Search completed: Sat Apr 15 00:53:50 2000  
Job time : 93 secs.

KW Nuclear protein; Phosphorylation; Anti-oncogene; DNA-binding;  
KW Transcription regulation; Activator.  
SQ SEQUENCE 393 AA; 43723 MW; DA7D302F CRC32;

Query Match 100.0%; Score 68; DB 4; Length 393;  
Best Local Similarity 100.0%; Pred. No. 2.57e-05;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 299 LPPGSTKRAL 308

QY 1 LPPGSTKRAL 10

RESULT 10  
ID Q16811 PRELIMINARY; PRT; 393 AA.

AC Q16811;  
DT 01-NOV-1996 (TREMBlrel. 01, Created)

DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)

DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)

DE CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

OC Eutheria; Primates; Catarrhini; Hominidae; Homo.

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE; 85126934.

RA MATLASHIEWSKI G., LAMB P., PIM D., PEACOCK J., CRAWFORD L.,

RA BENCHIMOL S.;

RT "Isolation and characterization of a human p53 cDNA clone: expression

of the human p53 gene.";

RL EMBO J. 3:3257-3262(1984).

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE; 87064416.

RA LAMB P., CRAWFORD L.;

RT "Characterization of the human p53 gene.";

RL Mol. Cell. Biol. 6:1379-1385(1986).

CC -1- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT

PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL

CYCLE REGULATION, IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY

REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED

FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF

CYCLIN-DEPENDENT KINASES (BY SIMILARITY).

CC -1- SUBCELLULAR LOCATION: NUCLEAR.

DR EMBL; M13121; AAA59987.1; JOINED.

DR EMBL; M13112; AAA59987.1; JOINED.

DR EMBL; M13113; AAA59987.1; JOINED.

DR EMBL; M13114; AAA59987.1; JOINED.

DR EMBL; M13115; AAA59987.1; JOINED.

DR EMBL; M13116; AAA59987.1; JOINED.

DR EMBL; M13117; AAA59987.1; JOINED.

DR EMBL; M13118; AAA59987.1; JOINED.

DR EMBL; M13119; AAA59987.1; JOINED.

DR EMBL; M13120; AAA59987.1; JOINED.

DR HSSP; P04637; LTSR.

DR PROSITE; PS00348; P53; 1.

DR PFAM; PF00870; P53; 1.

KW Repeat; Tumor antigen; Anti-oncogene; DNA-binding;

KW Transcription regulation; Activator; Nuclear protein; Phosphorylation.

FT NON\_TER 393

SQ SEQUENCE 393 AA; 43698 MW; 3EA71431 CRC32;

Query Match 100.0%; Score 68; DB 4; Length 393;

Best Local Similarity 100.0%; Pred. No. 2.57e-05;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 299 LPPGSTKRAL 308

QY 1 LPPGSTKRAL 10

RESULT 11

ID O70366 PRELIMINARY; PRT; 390 AA.

O70366;  
AC 01-AUG-1998 (TREMBlrel. 07, Created)  
DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)  
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)  
DE CELLULAR TUMOR ANTIGEN P53.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-LYMPHOID LEUKEMIA;  
RA PROTESJO L., NILSSON J., WANDZIOCH E., HEBY O.;  
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT  
PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL  
CYCLE REGULATION, IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY  
REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED  
FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF  
CYCLIN-DEPENDENT KINASES (BY SIMILARITY).

CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
DR EMBL; AF051368; AAC05704.1; -.  
DR HSSP; P04637; 1PET.  
DR PROSITE; PS00348; P53; 1.  
DR PFAM; PF00870; P53; 1.  
DR PRINTS; PR00386; P53SUPPRESSR.  
KW Anti-oncogene; DNA-binding; transcription regulation; Activator;  
KW Nuclear protein; Phosphorylation.  
SQ SEQUENCE 390 AA; 43430 MW; EDF4C8AA CRC32;

Db 296 LPPGSTKRAL 305

QY 1 LPPGSTKRAL 10

RESULT 12

ID Q29475 PRELIMINARY; PRT; 281 AA.

AC Q29475;  
DT 01-NOV-1996 (TREMBlrel. 01, Created)

DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)

DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)

DE CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).

GN P53.

OS Canis familiaris (Dog).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

OC Eutheria; Carnivora; Fissipedia; Canidae; Canis.

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE-MAMMARY GLAND;

RA MEDLINE; 97194812.

RA VAN LEEUWEN I., RUTTEMAN G.R., HELLMAN E., CORNELISSE C.C.J.,

RA DEVILLE P.;

RT "P53 mutations in mammary tumor cell lines and corresponding tumor

tissues in the dog.";

RL Anticancer Res. 16:3737-3744(1996).

CC -1- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT

PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL

CYCLE REGULATION, IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY

REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED

FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF

CYCLIN-DEPENDENT KINASES (BY SIMILARITY).

CC -1- SUBCELLULAR LOCATION: NUCLEAR.

DR EMBL; L37107; AAC37335.1; -.

DR HSSP; P04637; 1SAH.

DR PROSITE; PS00348; P53; 1.

DR PFAM; PF00870; P53; 1.

KW Anti-oncogene; DNA-binding; Transcription regulation; Activator;

KW Nuclear protein; Phosphorylation.

FT NON\_TER 1

FT NON\_TER 281

DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
 DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)  
 DE CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).  
 GN P53.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 92007731.  
 RA FARRELL P.J., ALLAN G., SHANAHAN F., VOUSDEN K.H., CROOK T.;  
 RT "p53 is frequently mutated in Burkitt's lymphoma cell lines."  
 RL EMBO J. 10:2879-2887(1991).  
 CC -1- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT  
 CC PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL  
 CC CYCLE REGULATION, IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY  
 CC REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED  
 CC FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF  
 CC CYCLIN-DEPENDENT KINASES (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 DR EMBL; X60018; CAA42633.1; -.  
 DR HSSP; P04637; 1SAH.  
 DR PROSITE; PS00348; P53; 1.  
 DR PFAM; PF00870; P53; 1.  
 DR Anti-oncogene; DNA-binding; Transcription regulation; Activator;  
 KW Nuclear protein; Phosphorylation.  
 FT VARIANT 163 163 H -> Y.  
 FT NON\_TER 393 393  
 SQ SEQUENCE 393 AA; 43627 MW; AFD8A9E3 CRC32;

Query Match 100.0%; Score 68; DB 4; Length 393;  
 Best Local Similarity 100.0%; Pred. No. 2.57e-05;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 299 LPPGSTRAL 308  
 |||||  
 QY 1 LPPGSTRAL 10

RESULT 7 PRELIMINARY; PRT; 393 AA.  
 ID Q16535  
 AC Q16535;  
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
 DT 01-NOV-1999 (TrEMBLrel. 01, Last sequence update)  
 DE P53 TRANSFORMATION SUPPRESSOR (FRAGMENT).  
 GN P53.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 92007731.  
 RA FARRELL P.J., ALLAN G., SHANAHAN F., VOUSDEN K.H., CROOK T.;  
 RT "p53 is frequently mutated in Burkitt's lymphoma cell lines."  
 RL EMBO J. 10:2879-2887(1991).  
 DR EMBL; X60017; CAA42632.1; -.  
 DR EMBL; X60015; CAA42630.1; -.  
 DR HSSP; P04637; 1SAH.  
 DR PFAM; PF00870; P53; 1.  
 FT VARIANT 248 248 Q -> R.  
 FT NON\_TER 393 393  
 SQ SEQUENCE 393 AA; 43684 MW; 239818A9 CRC32;

Query Match 100.0%; Score 68; DB 4; Length 393;  
 Best Local Similarity 100.0%; Pred. No. 2.57e-05;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 299 LPPGSTRAL 308  
 |||||  
 QY 1 LPPGSTRAL 10

RESULT 8 PRELIMINARY; PRT; 393 AA.  
 ID Q16809;  
 AC Q16809;  
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
 DT 01-NOV-1999 (TrEMBLrel. 01, Last sequence update)  
 DE CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).  
 GN P53.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 92007731.  
 RA FARRELL P.J., ALLAN G., SHANAHAN F., VOUSDEN K.H., CROOK T.;  
 RT "p53 is frequently mutated in Burkitt's lymphoma cell lines."  
 RL EMBO J. 10:2879-2887(1991).  
 CC -1- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT  
 CC PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL  
 CC CYCLE REGULATION, IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY  
 CC REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED  
 CC FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF  
 CC CYCLIN-DEPENDENT KINASES (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 DR EMBL; X60019; CAA42634.1; -.  
 DR HSSP; P04637; 1SAH.  
 DR PROSITE; PS00348; P53; 1.  
 DR PFAM; PF00870; P53; 1.  
 DR Anti-oncogene; DNA-binding; Transcription regulation; Activator;  
 KW Nuclear protein; Phosphorylation.  
 FT VARIANT 213 213 Q -> R.  
 FT NON\_TER 393 393  
 SQ SEQUENCE 393 AA; 43684 MW; CB70BD7F CRC32;

Query Match 100.0%; Score 68; DB 4; Length 393;  
 Best Local Similarity 100.0%; Pred. No. 2.57e-05;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 299 LPPGSTRAL 308  
 |||||  
 QY 1 LPPGSTRAL 10

RESULT 9 PRELIMINARY; PRT; 393 AA.  
 ID Q16848;  
 AC Q16848;  
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
 DT 01-NOV-1999 (TrEMBLrel. 01, Last sequence update)  
 DE CELLULAR TUMOR ANTIGEN P53.  
 GN TP53.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 87089826.  
 RA HARRIS N., BRILL E., SHOHAT O., PROKOCIMER M., WOLF D., ARAI N.,  
 RA ROTTER V.;  
 RT "Molecular basis for heterogeneity of the human p53 protein."  
 RL Mol. Cell. Biol. 6:4650-4656(1986).  
 CC -1- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT  
 CC PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL  
 CC CYCLE REGULATION, IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY  
 CC REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED  
 CC FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF  
 CC CYCLIN-DEPENDENT KINASES (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 DR EMBL; M14694; AAA61211.1; -.  
 DR HSSP; P04637; 1TSR.  
 DR PROSITE; PS00348; P53; 1.  
 DR PFAM; PF00870; P53; 1.  
 DR PRINTS; PR00386; P53SUPPRESSR.



```

RN RP SEQUENCE FROM N.A.
RX FARRELL P.J., ALLAN G., SHANAHAN F., VOUSDEN K.H., CROOK T.;
RA "p53 is frequently mutated in Burkitt's lymphoma cell lines.";
RT EMO J. 10:2879-2887(1991).
RL EMO J. 10:2879-2887(1991).
DR EMBL; X60016; CAA42631.1; -.
DR HSSP; P04637; ISAH.
DR PFAM; PF00870; P53; 1.
DR VARIANT 238 238 Y -> C.
FT NON_TER 393 393
SQ SEQUENCE 393 AA; 43713 MW; A01E1523 CRC32;

Query Match 100.0%; Score 68; DB 4; Length 393;
Best Local Similarity 100.0%; Pred. No. 2.57e-05;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 299 LPPGSTKRAL 308
QY 1 LPPGSTKRAL 10

RESULT 3
ID Q15086 PRELIMINARY; PRT; 393 AA.
AC Q15086;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
DE P53 TRANSFORMATION SUPPRESSOR (FRAGMENT).
GN P53.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 92007731.
RA FARRELL P.J., ALLAN G., SHANAHAN F., VOUSDEN K.H., CROOK T.;
RT "p53 is frequently mutated in Burkitt's lymphoma cell lines.";
RL EMO J. 10:2879-2887(1991).
DR EMBL; X60013; CAA42628.1; -.
DR HSSP; P04637; ISAH.
DR PFAM; PF00870; P53; 1.
DR VARIANT 246 246 T -> M.
FT NON_TER 393 393
SQ SEQUENCE 393 AA; 43682 MW; 943B62A3 CRC32;

Query Match 100.0%; Score 68; DB 4; Length 393;
Best Local Similarity 100.0%; Pred. No. 2.57e-05;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 299 LPPGSTKRAL 308
QY 1 LPPGSTKRAL 10

RESULT 4
ID Q16810 PRELIMINARY; PRT; 393 AA.
AC Q16810;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
DE CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).
GN P53.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 92007731.
RA FARRELL P.J., ALLAN G., SHANAHAN F., VOUSDEN K.H., CROOK T.;
RT "p53 is frequently mutated in Burkitt's lymphoma cell lines.";
RL EMO J. 10:2879-2887(1991).
DR EMBL; X60013; CAA42628.1; -.
DR HSSP; P04637; ISAH.
DR PFAM; PF00870; P53; 1.
DR VARIANT 246 246 T -> M.
FT NON_TER 393 393
SQ SEQUENCE 393 AA; 43682 MW; 943B62A3 CRC32;

Query Match 100.0%; Score 68; DB 4; Length 393;
Best Local Similarity 100.0%; Pred. No. 2.57e-05;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 299 LPPGSTKRAL 308
QY 1 LPPGSTKRAL 10

RESULT 5
ID Q16807 PRELIMINARY; PRT; 393 AA.
AC Q16807;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
DE CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).
GN P53.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 92007731.
RA FARRELL P.J., ALLAN G., SHANAHAN F., VOUSDEN K.H., CROOK T.;
RT "p53 is frequently mutated in Burkitt's lymphoma cell lines.";
RL EMO J. 10:2879-2887(1991).
DR EMBL; X60013; CAA42628.1; -.
DR HSSP; P04637; ISAH.
DR PFAM; PF00870; P53; 1.
DR VARIANT 246 246 T -> M.
FT NON_TER 393 393
SQ SEQUENCE 393 AA; 43714 MW; 5F914579 CRC32;

Query Match 100.0%; Score 68; DB 4; Length 393;
Best Local Similarity 100.0%; Pred. No. 2.57e-05;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 299 LPPGSTKRAL 308
QY 1 LPPGSTKRAL 10

RESULT 6
ID Q16808 PRELIMINARY; PRT; 393 AA.
AC Q16808;
DT 01-NOV-1996 (TREMBLrel. 01, Created)

```

(TM)

Result No.	Score	Query		Length	DB	ID	Description	Pred. No.
		Match						
1	68	100.0	393	4	Q15087	P53 TRANSFORMATION SUP	2.57e-05	
2	68	100.0	393	4	Q15088	P53 TRANSFORMATION SUP	2.57e-05	
3	68	100.0	393	4	Q15086	P53 TRANSFORMATION SUP	2.57e-05	
4	68	100.0	393	4	Q16810	CELLULAR TUMOR ANTIGEN	2.57e-05	
5	68	100.0	393	4	Q16807	CELLULAR TUMOR ANTIGEN	2.57e-05	
6	68	100.0	393	4	Q16808	CELLULAR TUMOR ANTIGEN	2.57e-05	
7	68	100.0	393	4	Q16535	P53 TRANSFORMATION SUP	2.57e-05	
8	68	100.0	393	4	Q16809	CELLULAR TUMOR ANTIGEN	2.57e-05	
9	68	100.0	393	4	Q16848	CELLULAR TUMOR ANTIGEN	2.57e-05	
10	68	100.0	393	4	Q16811	CELLULAR TUMOR ANTIGEN	2.57e-05	
11	65	95.0	393	4	Q70366	CELLULAR TUMOR ANTIGEN	1.85e-04	
12	60	88.2	281	6	Q29475	CELLULAR TUMOR ANTIGEN	4.52e-03	
13	56	82.4	45	4	Q39659	CELLULAR PHOSPHOTEL	5.29e-02	
14	55	80.9	648	4	Q00392	UV RADIATION RESISTANC	9.63e-02	
15	54	79.4	205	11	Q35873	CELLULAR TUMOR ANTIGEN	1.74e-01	
16	53	77.9	238	14	P99004	P53 (FRAGMENT) .	3.13e-01	
17	53	77.9	286	14	P90332	P53 (FRAGMENT) .	3.13e-01	
18	53	77.9	286	14	P90003	P53 (FRAGMENT) .	3.13e-01	
19	53	77.9	378	14	P98002	P53 (FRAGMENT) .	3.13e-01	
20	51	75.0	192	6	Q80078	EPITHELIAL MUCIN (FRAG	9.92e-01	

Query Match 75.0%; Score 51; DB 1; Length 676;  
Best Local Similarity 77.8%; Pred. No. 4.76e-01;  
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
Db 641 VPPGSTKRS 649  
QY :|||||:  
1 LPPGSTKRA 9

Search completed: Sat Apr 15 00:51:58 2000  
Job time : 43 secs.

CC CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN  
 CC TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION, IT IS A  
 CC TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION  
 CC BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF  
 CC THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.  
 CC APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF  
 CC BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2  
 CC EXPRESSION.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -1- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY  
 CC OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED  
 CC IN MANY TYPES OF CANCER.  
 CC -1- SIMILARITY: BELONGS TO THE P53 FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; M75144; AAA37085.1; -  
 CC EMBL; U07182; AAB41344.1; -  
 CC PIR; JH0633; JH0633.  
 CC HSP; P04637; LYCO.  
 CC PROSITE; PS00348; P53; 1.  
 CC PFAM; PF00870; P53; 1.  
 CC  
 CC Anti-oncogene; DNA-binding; Transcription regulation; Activator;  
 CC Nuclear protein; Phosphorylation; Apoptosis.  
 CC DOMAIN 1 77 ASP/GLU-RICH (ACIDIC).  
 CC DOMAIN 78 153 HYDROPHOBIC.  
 CC DOMAIN 319 393 HIGHLY BASIC AND MAY BE INVOLVED IN  
 CC INTERACTION WITH DNA.  
 CC DOMAIN 314 326 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 CC MOD\_RES 395 395 PHOSPHORYLATION (BY SIMILARITY).  
 CC CONFLICT 188 188 G -> S (IN REF. 2).  
 CC SEQUENCE 396 AA; 43631 MW; C3668ADE CRC32;  
 CC  
 CC Query Match 79.4%; Score 54; DB 1; Length 396;  
 CC Best Local Similarity 80.0%; Pred. No. 8.40e-02;  
 CC Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 CC  
 CC Db 302 LPPKSAKRAL 311  
 CC ||| |  
 CC QY 1 LPPGSTRKAL 10  
 CC  
 CC RESULT 14  
 CC ID DPYS\_RAT STANDARD; PRT; 519 AA.  
 CC AC Q63150;  
 CC DT 15-JUL-1998 (Rel. 36, Created)  
 CC DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 CC DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 CC DE DIHYDROXYRIMIDINASE (EC 3.5.2.2) (DHPASE) (HYDANTOINASE) (DHP).  
 CC GN DPYS.  
 CC OS Rattus norvegicus (Rat).  
 CC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 CC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 CC [1]  
 CC RN SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 CC RP TISSUE=LIVER;  
 CC RX MEDLINE; 96283806.  
 CC RA MATSUDA K., SAKATA S., KANEKO M., HAMAJIMA N., NONAKA M., SASAKI M.,  
 CC TAMAKI N.;  
 CC RT "Molecular cloning and sequencing of a cDNA encoding  
 CC dihydroxyrimidinase from the rat liver."  
 CC RL Biochim. Biophys. Acta 1307:140-144(1996).  
 CC -1- CATALYTIC ACTIVITY: 5.6-DIHYDROURACIL + H(2)O - 3-  
 CC UREIDOPROPIONATE.  
 CC -1- SIMILARITY: BELONGS TO THE DIHYDROXYRIMIDINASE FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; D63704; BAA09833.1; -  
 CC DR PFAM; PF00744; Dihydroorotase; 1.  
 CC KW Hydrolase.  
 CC SQ SEQUENCE 519 AA; 56833 MW; 631A9821 CRC32;  
 CC  
 CC Query Match 77.9%; Score 53; DB 1; Length 519;  
 CC Best Local Similarity 70.0%; Pred. No. 1.51e-01;  
 CC Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
 CC  
 CC Db 40 LPPGDTSRGL 49  
 CC |||| |  
 CC QY 1 LPPGSTRKAL 10  
 CC  
 CC RESULT 15  
 CC ID MUC1\_MESAU STANDARD; PRT; 676 AA.  
 CC AC Q60528;  
 CC DT 15-JUL-1999 (Rel. 38, Created)  
 CC DT 15-JUL-1999 (Rel. 38, Last sequence update)  
 CC DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 CC DE MUCIN 1 PRECURSOR.  
 CC GN MUC1.  
 CC OS Mesocricetus auratus (Golden hamster).  
 CC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 CC Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae; Mesocricetus.  
 CC [1]  
 CC RP SEQUENCE FROM N.A.  
 CC RC TISSUE=TRACHEAL EPITHELIUM;  
 CC RX MEDLINE; 96326118.  
 CC RT PARK H., HYUN S.W., KIM K.C.;  
 CC RT "Expression of MUC1 mucin gene by hamster tracheal surface epithelial  
 CC cells in primary culture."  
 CC RL Am. J. Respir. Cell Mol. Biol. 15:237-244(1996).  
 CC -1- FUNCTION: DIRECT OR INDIRECT INTERACTION WITH ACTIN  
 CC CYTOSKELETON (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.  
 CC -1- PTM: HIGHLY O-GLYCOSYLATED AND PROBABLY ALSO N-GLYCOSYLATED.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; U36918; AAB5965.1; -  
 CC DR PFAM; PF01390; SEA; 1.  
 CC KW Glycoprotein; Signal; Cytoskeleton; Actin-binding; Transmembrane;  
 CC Repeat.  
 CC FT SIGNAL 1 ? POTENTIAL.  
 CC FT CHAIN ? 676 MUCIN 1.  
 CC FT DOMAIN ? 582 EXTRACELLULAR (POTENTIAL).  
 CC FT TRANSMEM 583 603 POTENTIAL.  
 CC FT DOMAIN 604 676 CYTOPLASMIC (POTENTIAL).  
 CC FT CARBOHYD 291 291 POTENTIAL.  
 CC FT CARBOHYD 323 323 POTENTIAL.  
 CC FT CARBOHYD 350 350 POTENTIAL.  
 CC FT CARBOHYD 380 380 POTENTIAL.  
 CC FT CARBOHYD 400 400 POTENTIAL.  
 CC FT CARBOHYD 413 413 POTENTIAL.  
 CC FT CARBOHYD 435 435 POTENTIAL.  
 CC FT CARBOHYD 479 479 POTENTIAL.  
 CC FT CARBOHYD 496 496 POTENTIAL.  
 CC FT CARBOHYD 536 536 POTENTIAL.  
 CC SQ SEQUENCE 676 AA; 67616 MW; 5858458D CRC32;

TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES. APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2 EXPRESSION.

-1- SUBCELLULAR LOCATION: NUCLEAR.

-1- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED IN MANY TYPES OF CANCER.

-1- SIMILARITY: BELONGS TO THE P53 FAMILY.

THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; D26608; BAA05653.1; -  
 EMBL; D16460; BAA03927.1; -  
 DR HSP; P04637; 1SAH.  
 DR PROSITE; PS00348; P53; 1.  
 DR PFAM; PF00870; P53; 1.  
 KW Anti-oncogene; DNA-binding; Transcription regulation; Activator;  
 Nuclear protein; Phosphorylation; Apoptosis.  
 FT DOMAIN 1 59 ASP/GLU-RICH (ACIDIC).  
 FT MOD\_RES 304 316 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 FT MOD\_RES 385 385 PHOSPHORYLATION (BY SIMILARITY).  
 FT CONFLICT 285 285 K -> R (IN REF. 2).  
 SQ SEQUENCE 386 AA; 42692 MW; D6C7132A CRC32;

Query Match 88.2%; Score 60; DB 1; Length 386;  
 Best Local Similarity 100.0%; Pred. No. 2.19e-03;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 293 PPGSKRAL 301  
 |||||  
 QY 2 PPGSKRAL 10

RESULT 12  
 ID P53\_CRIGR STANDARD; PRT; 393 AA.  
 AC O09185; Q64397; P97258; P97788;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE CELLULAR TUMOR ANTIGEN P53.  
 GN TP53 OR P53.  
 OS Cricetus griseus (Chinese hamster).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae; Cricetulus.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA CHAUNG W., MI L.J., BOORSTEIN R.J.;  
 RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA TISSUE=LIVER;  
 RX MEDLINE; 97183659.  
 RA LEE H., LARNER J.M., HAMLIN J.L.;  
 RT "Cloning and characterization of Chinese hamster p53 cDNA."  
 RL Gene 184:177-183(1997).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 KC TISSUE=EMBRYONIC FIBROBLAST;  
 RA SHIMIZU T., NIKAI O., SUZUKI F.;  
 RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL CIRCUMSTANCES OR CELL TYPE. BUT BOTH ACTIVITIES ARE INVOLVED IN TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION, IT IS A

TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES. APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2 EXPRESSION.

-1- SUBCELLULAR LOCATION: NUCLEAR.

-1- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED IN MANY TYPES OF CANCER.

-1- SIMILARITY: BELONGS TO THE P53 FAMILY.

THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; Y08900; CAA70108.1; -  
 EMBL; Y08901; CAA70109.1; -  
 DR EMBL; U50395; AAC53040.1; -  
 DR EMBL; D86070; BAA13004.1; -  
 DR HSP; P04637; 1YCO.  
 DR PROSITE; PS00348; P53; 1.  
 DR PFAM; PF00870; P53; 1.  
 KW Anti-oncogene; DNA-binding; Transcription regulation; Activator;  
 Nuclear protein; Phosphorylation; Apoptosis.  
 FT DOMAIN 1 74 ASP/GLU-RICH (ACIDIC).  
 FT DOMAIN 75 150 HYDROPHOBIC.  
 FT DOMAIN 316 390 HIGHLY BASIC AND MAY BE INVOLVED IN INTERACTION WITH DNA.  
 FT DOMAIN 311 323 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 FT MOD\_RES 392 392 PHOSPHORYLATION (BY SIMILARITY).  
 FT VARIANT 133 133 L -> O (IN CELL LINE V79-4).  
 FT VARIANT 135 135 C -> W (IN CELL LINE V79-4).  
 FT CONFLICT 103 103 Y -> F (IN REF. 2).  
 SQ SEQUENCE 393 AA; 43378 MW; 402EB149 CRC32;

Query Match 79.4%; Score 54; DB 1; Length 393;  
 Best Local Similarity 80.0%; Pred. No. 8.40e-02;  
 Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 299 LPPSKRAL 308  
 |||||  
 QY 1 LPPGSKRAL 10

RESULT 13  
 ID P53\_MESAU STANDARD; PRT; 396 AA.  
 AC Q00366; P97276;  
 DT 01-DEC-1992 (Rel. 24, Created)  
 DT 01-DEC-1992 (Rel. 24, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE CELLULAR TUMOR ANTIGEN P53.  
 GN TP53.  
 OS Mesocricetus auratus (Golden hamster).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae; Mesocricetus.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=SYRIAN; TISSUE=KIDNEY;  
 RX MEDLINE; 92210007.  
 RA LEGROS Y., MCINTYRE P., SOUSSI T.;  
 RT "The cDNA cloning and immunological characterization of hamster p53."  
 RL Gene 112:247-250(1992).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA HOU E.W., WISEMAN R.;  
 RL Submitted (APR-1994) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL

```

CC STRAIN-BEAGLE;
CC MEDLINE; 95323915.
CC
CC RA KRAEGL S.A., PAZZI K.A., MADEWELL B.R.;
CC "Sequence analysis of canine p53 in the region of exons 3-8.";
CC Cancer Lett. 92:181-186(1995).
CC
CC CC -1- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES
CC GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL
CC CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN
CC TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION
CC BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF
CC THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.
CC APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF
CC BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2
CC EXPRESSION.
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- SIMILARITY: BELONGS TO THE P53 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC DR EMBL; X81705; CAA57349.1; -
CC DR HSSP; P04637; 1PEP.
CC DR PROSITE; PS00348; P53; 1.
CC DR PFAM; PF00870; P53; 1.
CC KW Anti-oncogene; DNA-binding; Transcription regulation; Activator;
CC Nuclear protein; Phosphorylation; Apoptosis.
CC FT DOMAIN 1 55 ASP/GLU-RICH (ACIDIC)
CC FT DOMAIN 300 322 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
CC FT MOD_RES 381 381 PHOSPHORYLATION (BY SIMILARITY).
CC SQ SEQUENCE 382 AA; 42809 MW; 0CB99A00 CRC32;
CC
CC Query Match 88.2%; Score 60; DB 1; Length 382;
CC Best Local Similarity 100.0%; Pred. No. 2.19e-03;
CC Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CC
CC DB 289 PPGSTKRAL 297
CC QY 2 PPGSTKRAL 10
CC
CC RESULT 11
CC ID P53_FELCA STANDARD; PRT; 386 AA.
CC AC P41685;
CC DT 01-NOV-1995 (Rel. 32, Created)
CC DT 01-NOV-1995 (Rel. 32, Last sequence update)
CC DT 01-NOV-1997 (Rel. 35, Last annotation update)
CC DE CELLULAR TUMOR ANTIGEN P53.
CC GN TP53.
CC OS Felis silvestris catus (Cat).
CC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
CC OC Eutheria; Carnivora; Fissipedia; Felidae; Felis.
CC RN [1]
CC RP SEQUENCE FROM N.A.
CC RC TISSUE=LYMPH NODE;
CC RX MEDLINE; 94333960.
CC RA OKUDA M., UEDA A., SAKAI T., OHASHI T., MOMOI Y., YOUN H.Y.,
CC WATARI T., GOITSUKA R., TSUJIMOTO H., HASEGAWA A.;
CC "Cloning of feline p53 tumor-suppressor gene and its aberration in
CC RT hematopoietic tumors.";
CC RN Int. J. Cancer 58:602-607(1994).
CC RN [2]
CC RP SEQUENCE OF 34-354 FROM N.A.
CC RX MEDLINE; 94114699.
CC RA OKUDA M., UEDA A., MATSUMOTO Y., MOMOI Y., WATARI T., GOITSUKA R.,
CC O'BRIEN S.J., TSUJIMOTO H., HASEGAWA A.;
CC "Molecular cloning and chromosomal mapping of feline p53 tumor
CC RT suppressor gene.";
CC RN J. Vet. Med. Sci. 55:801-805(1993).
CC CC -1- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES
CC GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL
CC CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN
CC TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION, IT IS A

```

RA MATHUPALA S.P.:  
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES  
CC GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL  
CC CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN  
CC TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION, IT IS A  
CC TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION  
CC BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF  
CC THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.  
CC APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF  
CC BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2  
CC EXPRESSION.  
CC -!- SUBCELLULAR LOCATION: NUCLEAR.  
CC -!- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY  
CC OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED  
CC IN MANY TYPES OF CANCER.  
CC -!- SIMILARITY: BELONGS TO THE P53 FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; X13058; CAA31457.1; -  
DR EMBL; L07910; AAA41788.1; JOINED.  
DR EMBL; L07904; AAA41788.1; JOINED.  
DR EMBL; L07905; AAA41788.1; JOINED.  
DR EMBL; L07906; AAA41788.1; JOINED.  
DR EMBL; L07907; AAA41788.1; JOINED.  
DR EMBL; L07908; AAA41788.1; JOINED.  
DR EMBL; L07909; AAA41788.1; JOINED.  
DR EMBL; U90328; AAB80959.1; -  
DR PIR; S02192; S02192.  
DR HSSP; P04637; IPET.  
DR PFAM; PS00348; P53; 1.  
DR PFAM; PF00870; P53; 1.  
KW Anti-oncogene; DNA-binding; Transcription regulation; Activator;  
KW Nuclear protein; Phosphorylation.  
FT DOMAIN 1 76  
FT DOMAIN 77 151 ASP/GLU-RICH (ACIDIC).  
FT DOMAIN 277 391 HYDROPHOBIC.  
FT DOMAIN 309 321 HIGHLY BASIC AND MAY BE INVOLVED IN  
FT MOD\_RES 390 390 INTERACTION WITH DNA.  
FT VARIANT 103 103 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
FT VARIANT 256 256 PHOSPHORYLATION (BY SIMILARITY).  
FT CONFLICT 174 174 G -> S.  
FT CONFLICT 174 174 E -> G.  
FT CONFLICT 174 174 C -> W (IN REF. 2).  
SQ SEQUENCE 391 AA; 43451 MW; E0114C18 CRC32;  
  
Query Match 95.68; Score 65; DB 1; Length 391;  
Best Local Similarity 90.08; Pred. No. 8.93e-05;  
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
  
Db 297 LPPGSAKRAL 306  
YQ 1 LPPGSAKRAL 10  
| | | | | | | | | |  
  
RESULT 8  
ID P53\_SPEBE STANDARD; PRT; 314 AA.  
AC Q64662;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).  
GN TP53.  
OS Spermophilus beecheyi (Beechey ground squirrel).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;  
OC Eutheria; Rodentia; Sciurognathi; Scluridae; Sclurinae; Spermophilus.  
RN [1]

RP SEQUENCE FROM N.A.  
RC TISSUE=THYMUS;  
RX MEDLINE; 95007566.  
RA RIVKINA M.B.; CULLEN J.M.; ROBINSON W.S.; MARION P.L.;  
RT "State of the p53 gene in hepatocellular carcinomas of ground  
RT squirrels and woodchucks with past and ongoing infection with  
RT hepadnaviruses.";  
RL Cancer Res. 54:5430-5437(1994).  
CC -!- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES  
CC GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL  
CC CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN  
CC TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION, IT IS A  
CC TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION  
CC BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF  
CC THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.  
CC APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF  
CC BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2  
CC EXPRESSION.  
CC -!- SUBCELLULAR LOCATION: NUCLEAR.  
CC -!- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY  
CC OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED  
CC IN MANY TYPES OF CANCER.  
CC -!- SIMILARITY: BELONGS TO THE P53 FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; U43902; AAA85628.1; -  
DR HSSP; P04637; 1YCS.  
DR PROSITE; PS00348; P53; 1.  
DR PFAM; PF00870; P53; 1.  
KW Anti-oncogene; DNA-binding; Transcription regulation; Activator;  
KW Nuclear protein; Phosphorylation.  
FT NON\_TER 1  
FT DOMAIN 289 301  
FT NON\_TER 314 314 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
SQ SEQUENCE 314 AA; 34618 MW; D07F433B CRC32;  
  
Query Match 88.24; Score 60; DB 1; Length 314;  
Best Local Similarity 100.0%; Pred. No. 2.19e-03;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Db 278 PPGSTKRAL 286  
YQ 2 PPGSTKRAL 10  
| | | | | | | | | |  
  
RESULT 9  
ID P53\_CANFA STANDARD; PRT; 381 AA.  
AC Q29337;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 15-DEC-1998 (Rel. 37, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE CELLULAR TUMOR ANTIGEN P53.  
GN TP53 OR P53.  
OS Canis familiaris (Dog).  
OC Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Mammalia;  
OC Eutheria; Carnivora; Fissipedia; Canidae; Canis.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=LEUKOCYTE;  
RX MEDLINE; 98178696.  
RA VELDHOEN N.; MILNER J.;  
RT "Isolation of canine p53 cDNA and detailed characterization of the  
RT full length canine p53 protein.";  
RL Oncogene 16:1077-1084(1998).  
RN [2]  
RP SEQUENCE OF 25-300 FROM N.A.

OC [1]  
RN SEQUENCE FROM N.A.  
RP  
RX MEDLINE: 89083585.  
RA  
RA SOUSSI T.;  
RT "Nucleotide sequence of a cDNA encoding the rat p53 nuclear  
RT oncoprotein.";  
RL Nucleic Acids Res. 16:11384-11384(1988).  
RN [2]  
RN SEQUENCE FROM N.A.  
RP  
RX MEDLINE: 93181368.  
RA HULLA J.E., SCHNEIDER R.P.;  
RT "Structure of the rat p53 tumor suppressor gene.";  
RL Nucleic Acids Res. 21:713-717(1993).  
RN [3]  
RN SEQUENCE FROM N.A.  
RP  
RC STRAIN-SPRAGUE-DAWLEY;



```
RESULT 4
ID P53_CERAE STANDARD; PRT; 393 AA.
AC P13481;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DE 01-NOV-1997 (Rel. 35, Last annotation update)
DE CELLULAR TUMOR ANTIGEN P53.
GN TP53.
OS Cercopithecus aethiops (Green monkey) (Grivet).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Cercopithecoidea; Cercopithecinae;
OC Chlorocebus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 85027173.
RA BTENZ B., ZAKUT-HOURI R., GIVOL D., OREN M.;
RT "Analysis of the gene coding for the murine cellular tumour antigen
p53.";
RL EMBO J. 3:2179-2183(1984).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 84068204.
RA ZAKUT-HOURI R., OREN M., BIENZ B., LAVIE V., HAZUM S., GIVOL D.;
RT "A single gene and a pseudogene for the cellular tumour antigen p53.";
RL Nature 306:594-597(1983).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE; 84272240.
RA JENKINS J.R., RUDGE K., REDMOND S., WADE-EVANS A.;
RT "Cloning and expression analysis of full length mouse cDNA sequences
encoding the transformation associated protein p53.";
RL Nucleic Acids Res. 12:5609-5626(1984).
RN [4]
RP SEQUENCE FROM N.A. (CLONES PCD53; P53-M11 AND P53-M8).
RX MEDLINE; 87064640.
RA ARAI N., NOMURA D., YOKOTA K., WOLF D., BRILL E., SHOHAT O.,
RT "Immunologically distinct p53 molecules generated by alternative
splicing.";
RL Mol. Cell. Biol. 6:3232-3239(1986).
RN [5]
RP SEQUENCE OF 222-258 FROM N.A.
RX MEDLINE; 92115342.
RA BURNS P.A., KEMP C.J., GANNON J.V., LANE D.P., BRENNER R.,
RA BALMAIN A.;
RT "Loss of heterozygosity and mutational alterations of the p53 gene in
skin tumours of interspecific hybrid mice.";
RL Oncogene 6:2363-2369(1991).
RN [6]
RP PHOSPHORYLATION SITES.
RX MEDLINE; 86149247.
RA SAMAD A., ANDERSON C.W., CARROLL R.B.;
RT "Mapping of phosphonoester and apparent phosphodiester bonds of the
oncogene product p53 from simian virus 40-transformed 3T3 cells.";
RL Proc. Natl. Acad. Sci. U.S.A. 83:897-901(1986).
RN [7]
RP PHOSPHORYLATION SITES.
RX MEDLINE; 91006019.
RA MEEK D.W., SIMON S., KIKKAWA U., ECKHART W.;
RT "The p53 tumour suppressor protein is phosphorylated at serine 389 by
casein kinase II.";
RL EMBO J. 9:3253-3260(1990).
CC -1- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES
GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL
CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN
TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION, IT IS A
TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION
BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF
THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.
APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF
BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2
EXPRESSION.
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY
OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED
IN MANY TYPES OF CANCER.
CC -1- SIMILARITY: BELONGS TO THE P53 FAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; X16384; CAA34420.1; -
DR PIR; S06594; S06594.
DR HSP; P04637; 1SAH.
DR PROSITE; PS00348; P53; 1.
DR PFAM; PF00870; P53; 1.
KW Anti-oncogene; DNA-binding; Transcription regulation; Activator;
FT DOMAIN 1 68 ASP/GLU-RICH (ACIDIC).
FT DOMAIN 81 150 HYDROPHOBIC.
FT DOMAIN 319 393 HIGHLY BASIC AND MAY BE INVOLVED IN
INTERACTION WITH DNA.
FT DOMAIN 311 323 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT MOD_RES 392 392 PHOSPHORYLATION (BY SIMILARITY).
SQ SEQUENCE 393 AA; 43696 MW; BBE7DC62 CRC32;
Query Match 100.0%; Score 68; DB 1; Length 393;
Best Local Similarity 100.0%; Pred. No. 1.23e-05;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 299 LPPGSTKRAL 308
QY 1 LPPGSTKRAL 10
RESULT 5
ID P53_MOUSE STANDARD; PRT; 390 AA.
AC P02340;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
```

RT "Genetic and immunochemical analysis of mutant p53 in human breast  
cancer cell lines."  
RL Oncogene 5:893-899(1990).  
RN [26]  
RP VARIANTS COLON TUMORS PHE-241 AND HIS-273.  
RX MEDLINE: 91017544.  
RA RODRIGUES N.R., ROWAN A., SMITH M.E.F., KERR I.B., BODMER W.F.,  
R GANNON J.V., LANE D.P.;

...  
Note: remainder of annotations omitted.

Query Match 100.0%; Score 68; DB 1; Length 393;  
Best Local Similarity 100.0%; Pred. No. 1.23e-05;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 299 LPPGSTKRAL 308  
| | | | | | | | | |  
QY 1 LPPGSTKRAL 10

RESULT 2  
ID P53\_MACFA STANDARD; PRT; 393 AA.  
AC P56423;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 15-JUL-1998 (Rel. 36, Last annotation update)  
DE CELLULAR TUMOR ANTIGEN P53.  
GN TP53 OR P53.  
OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Primates; Catarrhini; Cercopithecoidea; Cercopithecoinae;  
OC Macaca.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA KHAN M.A., HANSEN C., WELSH J.A., BENNETT W.P.;  
RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES  
CC GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL  
CC CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN  
CC TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION, IT IS A  
CC TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION  
CC BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF  
CC THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.  
CC APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF  
CC BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2  
CC EXPRESSION.  
CC -!- SUBCELLULAR LOCATION: NUCLEAR.  
CC -!- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY  
CC OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED  
CC IN MANY TYPES OF CANCER.

CC -!- SIMILARITY: BELONGS TO THE P53 FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; U48957; AAB91535.1; -;  
CC HSSP; P04637; 1SAH.  
CC PROSITE; PS00348; P53; 1.  
CC PFAM; PF00870; P53; 1.  
CC KW Anti-oncogene; DNA-binding; Transcription regulation; Activator;  
CC Nuclear protein; Phosphorylation; Apoptosis.  
CC FT DOMAIN 1 80 ASP/GLU-RICH (ACIDIC).  
CC FT DOMAIN 81 150 HYDROPHOBIC.  
CC FT DOMAIN 319 393 HIGHLY BASIC AND MAY BE INVOLVED IN  
CC INTERACTION WITH DNA.  
CC FT DOMAIN 311 323 NUCLEAR LOCALIZATION SIGNAL.  
CC FT MOD\_RES 392 392 PHOSPHORYLATION (BY SIMILARITY).  
CC SEQUENCE 393 AA; 43678 MW; 2499AC47 CRC32;

Query Match 100.0%; Score 68; DB 1; Length 393;  
Best Local Similarity 100.0%; Pred. No. 1.23e-05;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 299 LPPGSTKRAL 308  
| | | | | | | | | |  
QY 1 LPPGSTKRAL 10

Query Match 100.0%; Score 68; DB 1; Length 393;  
Best Local Similarity 100.0%; Pred. No. 1.23e-05;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 299 LPPGSTKRAL 308  
| | | | | | | | | |  
QY 1 LPPGSTKRAL 10

## RESULT 3

ID P53\_MACMU STANDARD; PRT; 393 AA.  
AC P56424;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE CELLULAR TUMOR ANTIGEN P53.  
GN TP53 OR P53.  
OS Macaca mulatta (Rhesus macaque).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Primates; Catarrhini; Cercopithecoidea; Cercopithecoinae;  
OC Macaca.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA KHAN M.A., HANSEN C., WELSH J.A., BENNETT W.P.;  
RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES  
CC GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL  
CC CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN  
CC TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION, IT IS A  
CC TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION  
CC BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF  
CC THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.  
CC APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF  
CC BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2  
CC EXPRESSION.  
CC -!- SUBCELLULAR LOCATION: NUCLEAR.  
CC -!- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY  
CC OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED  
CC IN MANY TYPES OF CANCER.

CC -!- SIMILARITY: BELONGS TO THE P53 FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; U48956; AAB91534.1; -;  
CC HSSP; P04637; 1SAH.  
CC PROSITE; PS00348; P53; 1.  
CC PFAM; PF00870; P53; 1.  
CC KW Anti-oncogene; DNA-binding; Transcription regulation; Activator;  
CC Nuclear protein; Phosphorylation; Apoptosis.  
CC FT DOMAIN 1 80 ASP/GLU-RICH (ACIDIC).  
CC FT DOMAIN 81 150 HYDROPHOBIC.  
CC FT DOMAIN 319 393 HIGHLY BASIC AND MAY BE INVOLVED IN  
CC INTERACTION WITH DNA.  
CC FT DOMAIN 311 323 NUCLEAR LOCALIZATION SIGNAL.  
CC FT MOD\_RES 392 392 PHOSPHORYLATION (BY SIMILARITY).  
CC SEQUENCE 393 AA; 43655 MW; 11A9B7F8 CRC32;

Query Match 100.0%; Score 68; DB 1; Length 393;  
Best Local Similarity 100.0%; Pred. No. 1.23e-05;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 299 LPPGSTKRAL 308  
| | | | | | | | | |  
QY 1 LPPGSTKRAL 10

RP SEQUENCE OF 101-393 FROM N.A.  
RX MEDLINE; 85126934.  
RA NATLASHESKI G., LAMB P., PIM D., PEACOCK J., CRAWFORD L.,  
RA BENCHIMOL S.;  
RT "Isolation and characterization of a human p53 cDNA clone: expression  
RT of the human p53 gene.";  
RL EMBO J. 3:3257-3262(1984).  
RN [7]  
RP NUCLEAR LOCALIZATION SIGNAL.  
RX MEDLINE; 90191730.  
RA ADDISON C., JENKINS J.R., STURZBECHER H.-W.;  
RT "The p53 nuclear localisation signal is structurally linked to a  
RT p34cdc2 kinase motif.";  
RL Oncogene 5:423-426(1990).  
RN [8]  
RP PHOSPHORYLATION BY P60/CDC2 AND CYCLIN B/CDC2.  
RX MEDLINE; 90280456.  
RA BISCHOFF J.R., FRIEDMAN P.N., MARSHAK D.R., PRIVES C., BEACH D.;  
RT "Human p53 is phosphorylated by p60-cdc2 and cyclin B-cdc2.";  
RL Proc. Natl. Acad. Sci. U.S.A. 87:4766-4770(1990).  
RN [9]  
RP DEPHOSPHORYLATION BY PP2A.  
RX MEDLINE; 91172186.  
RA SCHEIDTMANN K.H., MURPHY M.C., RUNDELL K., WALTER G.;  
RT "Dephosphorylation of simian virus 40 large-T antigen and p53 protein  
RT by protein phosphatase 2A: inhibition by small-t antigen.";  
RL Mol. Cell. Biol. 11:1996-2003(1991).  
RN [10]  
RP STRUCTURE BY NMR OF 319-360.  
RX MEDLINE; 94294808.  
RA CLORE G.M., OMICHINSKI J.G., SAKAGUCHI K., ZAMBRANO N., SAKAMOTO H.,  
RA APPELLA E., GRONENBORN A.M.;  
RT "High-resolution structure of the oligomerization domain of p53 by  
RT multidimensional NMR.";  
RL Science 265:386-391(1994).  
RN [11]  
RP STRUCTURE BY NMR OF 325-355.  
RX MEDLINE; 95292092.  
RA LEE W., HARVEY T.S., YIN Y., YAU P., LITCHFIELD D., ARROWSMITH C.H.;  
RT "Solution structure of the tetrameric minimum transforming domain of  
RT p53.";  
RL Nat. Struct. Biol. 1:877-890(1994).  
RN [12]  
RP STRUCTURE BY NMR OF 326-354.  
RX MEDLINE; 98026899.  
RA MCCOY M., STAVRIDIS E.S., WATERMAN J.L., WIECZOREK A.M., OPELLA S.J.,  
RA HALAZONETIS T.D.;  
RT "Hydrophobic side-chain size is a determinant of the  
RT three-dimensional structure of the p53 oligomerization domain.";  
RL EMBO J. 16:6230-6236(1997).  
RN [13]  
RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 94-289.  
RX MEDLINE; 94294806.  
RA CHO Y., GORINA S., JEFFREY P.D., PAVLETICH N.P.;  
RT "Crystal structure of a p53 tumor suppressor-DNA complex:  
RT understanding tumorigenic mutations.";  
RL Science 265:346-355(1994).  
RN [14]  
RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS) OF 13-29 IN COMPLEX WITH MDM2.  
RX MEDLINE; 97081050.  
RA KUSIIE P.H., GORINA S., MARECHAL V., ELENEAS B., MOREAU J.,  
RA LEVINE A.J., PAVLETICH N.P.;  
RT "Structure of the MDM2 oncoprotein bound to the p53 tumor suppressor  
RT transactivation domain.";  
RL Science 274:948-953(1996).  
RN [15]  
RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 97-287 IN COMPLEX WITH 53BP2.  
RX MEDLINE; 97035414.  
RA GORINA S., PAVLETICH N.P.;  
RT "Structure of the p53 tumor suppressor bound to the ankyrin and SH3  
RT domains of 53BP2.";  
RL Science 274:1001-1005(1996).  
RN [16]

RP REVIEW.  
RX MEDLINE; 94090335.  
RA HARRIS C.C.;  
RT "p53: at the crossroads of molecular carcinogenesis and risk  
RT assessment.";  
RL Science 262:1980-1981(1993).  
RN [17]  
RP REVIEW ON VARIANTS.  
RX MEDLINE; 91289156.  
RA HOOLSTEIN M., SIDRANSKY D., VOGELSTEIN B., HARRIS C.C.;  
RT "p53 mutations in human cancers.";  
RL Science 253:49-53(1991).  
RN [18]  
RP REVIEW ON VARIANTS.  
RX MEDLINE; 96271983.  
RA DE VRIES E.M.G., RICKE D.O., DE VRIES T.N., HARTMANN A., BLASZYK H.,  
RA LIAO D., SOUSSI T., KOVACH J.S., SOMMER S.S.;  
RT "Database of mutations in the p53 and APC tumor suppressor genes  
RT designed to facilitate molecular epidemiological analyses.";  
RL Hum. Mutat. 7:202-213(1996).  
RN [19]  
RP VARIANT ARG-72.  
RX MEDLINE; 91153807.  
RA OLSCHWANG S., LAURENT-PUIG P., VASSAL A., SALMON R.-J., THOMAS G.;  
RT "Characterization of a frequent polymorphism in the coding sequence  
RT of the Tp53 gene in colonic cancer patients and a control  
RT population.";  
RL Hum. Genet. 86:369-370(1991).  
RN [20]  
RP VARIANT LFS THR-133.  
RX MEDLINE; 92034774.  
RA LAW J.C., STRONG L.C., CHIDAMBARAM A., FERRELL R.E.;  
RT "A germ line mutation in exon 5 of the p53 gene in an extended cancer  
RT family.";  
RL Cancer Res. 51:6385-6387(1991).  
RN [21]  
RP VARIANTS LFS CYS-245; TRP-248; PRO-252 AND LYS-258.  
RX MEDLINE; 91057657.  
RA MALKIN D., LI F.P., STRONG L.C., FRAUMENI J.F. JR., NELSON C.E.,  
RA KIM D.H., KASSEL J., GRZYKA M.A., BISCHOFF F.Z., TAINSKY M.A.,  
RA FRIEND S.H.;  
RT "Germ line p53 mutations in a familial syndrome of breast cancer,  
RT sarcomas, and other neoplasms.";  
RL Science 250:1233-1238(1990).  
RN [22]  
RP VARIANT LFS ASP-245.  
RX MEDLINE; 91080929.  
RA SRIVASTAVA S., ZOU Z., PIROLLO K., BLATTNER W., CHANG E.H.;  
RT "Germ-line transmission of a mutated p53 gene in a cancer-prone  
RT family with Li-Fraumeni syndrome.";  
RL Nature 348:747-749(1990).  
RN [23]  
RP VARIANT LFS LEU-272.  
RX MEDLINE; 92147883.  
RA FELIX C.A., NAU M.M., TAKAHASHI T., MITSUDOMI T., CHIBA I.,  
RA POPLACK D.G., REAMAN G.H., COLE D.E., LETTERIO J.J., WHANG-PENG J.,  
RA KNUDSEN T., MINNA J.D.;  
RT "Hereditary and acquired p53 gene mutations in childhood acute  
RT lymphoblastic leukemia.";  
RL J. Clin. Invest. 89:640-647(1992).  
RN [24]  
RP VARIANTS LFS HIS-273 AND VAL-325.  
RX MEDLINE; 92228023.  
RA MALKIN D., JOLLY K.W., BARBIER N., LOOK A.T., FRIEND S.H.,  
RA GEBHART M.C., ANDERSEN T.I., BORRESEN A.-L., LI F.P., GARBER J.,  
RA STRONG L.C.;  
RT "Germline mutations of the p53 tumor-suppressor gene in children and  
RT young adults with second malignant neoplasms.";  
RL New Engl. J. Med. 326:1309-1315(1992).  
RN [25]  
RP VARIANTS BREAST TUMORS GLN-132; SER-249; LYS-280 AND LYS-285.  
RX MEDLINE; 90295284.  
RA BARTEK J., IGGO R., GANNON J., LANE D.P.;



**THIS PAGE BLANK (USPTO)**

```
744-761      #domain transmembrane #status predicted #label TM5\
760-796      #domain transmembrane #status predicted #label TM6\
874-897      #domain transmembrane #status predicted #label TM7\
902-924      #domain transmembrane #status predicted #label TM8\
SUMMARY      #length 963 #molecular-weight 109881 #checksum 1849

Query Match  72.1%; Score 49; DB 2; Length 963;
Best Local Similarity 70.0%; Pred. No. 3.89e+00;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 227 LPPGSKKAL 236
QY 1 LPPGSKKAL 10

RESULT 14
ENTRY      A41939      #type complete
TITLE      G protein-coupled glutamate receptor - rat
ORGANISM   G protein-coupled glutamate receptor #common_name Norway rat
DATE       04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change
           20-Mar-1998
ACCESSIONS A41939; S15362
REFERENCE   A41939
#authors   Houamed, K.M.; Kuijper, J.L.; Gilbert, T.L.; Haldeman, B.A.;
           O'Hara, P.J.; Mulvihill, E.R.; Almers, W.; Hagen, F.S.
#journal   Science (1991) 252:1318-1321
#title     Cloning, expression, and gene structure of a G
           protein-coupled glutamate receptor from rat brain.
#cross-references MUID:92022526
#accession A41939
##status   preliminary; not compared with conceptual translation
##molecule_type nucleic acid
##residues 1-1199 #label HOU
##cross-references GB:M61099; NID:g397806; PID:g204460
##experimental_source cerebellum
##note     S15362
REFERENCE   S15362
#authors   Masu, M.; Tanabe, Y.; Tsuchida, K.; Shigemoto, R.; Nakanishi,
           S.
#journal   Nature (1991) 349:760-765
#title     Sequence and expression of a metabotropic glutamate receptor.
#cross-references MUID:91156047
#accession S15362
##status   preliminary
##molecule_type mRNA
##residues 1-1199 #label MAS
##cross-references EMBL:X57569; NID:g56646; PID:g56647
KEYWORDS   G protein-coupled receptor; transmembrane protein
SUMMARY    #length 1199 #molecular-weight 133235 #checksum 5211

Query Match  72.1%; Score 49; DB 2; Length 1199;
Best Local Similarity 60.0%; Pred. No. 3.89e+00;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 147 LPPGRTKPI 156
QY 1 LPPGRTKPI 10

RESULT 15
ENTRY      S60200      #type complete
TITLE      acetyl-CoA carboxylase (EC 6.4.1.2) - smut fungus (Ustilago
           maydis)
ORGANISM   #formal_name Ustilago maydis #common_name corn smut
DATE       15-Feb-1996 #sequence_revision 01-Mar-1996 #text_change
           09-Sep-1997
ACCESSIONS S60200; S49991
REFERENCE   S60200
#authors   Bailey, A.; Keon, J.; Owen, J.; Hargreaves, J.
#journal   Mol. Gen. Genet. (1995) 249:191-201
#title     The ACC1 gene, encoding acetyl-CoA carboxylase, is essential
           for growth in Ustilago maydis.
#accession S60200
```

```
##molecule_type DNA
##residues 1-2185 #label BAI
##cross-references EMBL:Z46886; NID:g600037; PID:g600098
GENETICS
#gene      ACC1
#introns   14/1
CLASSIFICATION #superfamily lipoyl/biotin-binding homology; biotin
               carboxylase homology
KEYWORDS      biotin; ligase
FEATURE
41-548        #domain biotin carboxylase homology #label BCH
675-747       #domain lipoyl/biotin-binding homology #label LPS
714           #binding_site biotin (lys) (covalent) #status predicted
SUMMARY       #length 2185 #molecular-weight 240029 #checksum 9283

Query Match  72.1%; Score 49; DB 2; Length 2185;
Best Local Similarity 66.7%; Pred. No. 3.89e+00;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 155 PPGSAMRSL 163
QY 2 PPGSTKRAL 10

Search completed: Sat Apr 15 00:50:58 2000
Job time : 19 secs.
```

```

TITLE      episialin - mouse
ORGANISM   #formal_name Mus musculus #common_name house mouse
DATE       02-Aug-1996 #sequence_revision 02-Aug-1996 #text_change
REFERENCE  25-Oct-1996
ACCESSIONS 152257; I65210
#authors   Vos, H.L.; De Vries, Y.; Hilkens, J.
#journal   Biochem. Biophys. Res. Commun. (1991) 181:121-130
#title     The mouse episialin (Muc1) gene and its promoter. Rapid
           evolution of the repetitive domain in the protein.
#cross-references MUID:92068178
#accession 152257
#status    preliminary; translated from GB/EMBL/DBJ
#molecule_type DNA
#residues  1-631 #label RES
#cross-references GB:M77226; NID:g199835; PID:g199837
#accession 165210
#status    preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues  1-631 #label RE2
#cross-references GB:M84683; NID:g199842; PID:g199843
GENETICS   Muc1
#gene      20/1; 454/3; 472/2; 517/1; 557/3; 607/3
#introns   #length 631 #molecular-weight 54636 #checksum 6763
SUMMARY    Query Match 75.0%; Score 51; DB 2; Length 631;
           Best Local Similarity 77.8%; Pred. NO. 1.41e+00;
           Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 595 VPPGSTKRS 603
   :|||||:
QY 1 LPPGSTKRA 9

RESULT 11
ENTRY  I56979 #type fragment
TITLE  nitric-oxide synthase (EC 1.14.13.39) - rat (fragment)
ORGANISM #formal_name Rattus norvegicus #common_name Norway rat
DATE     26-Jul-1996 #sequence_revision 26-Jul-1996 #text_change
        24-Jul-1998
ACCESSIONS I56979
REFERENCE  Mohnhaupt, M.G.; Elzie, J.L.; Ahn, K.Y.; Clapp, W.L.; Wilcox,
           C.S.; Kone, B.C.
           Kidney Int. (1994) 46:653-665
           Differential expression and induction of mRNAs encoding two
           inducible nitric oxide synthases in rat kidney.
#cross-references MUID:95089280
#accession I56979
#status    preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues  1-230 #label RES
#cross-references EMBL:U02534; NID:g408464; PID:g408465
CLASSIFICATION #superfamily nitric-oxide synthase; flavodoxin homology;
           NADPH-ferrithemoprotein reductase homology
KEYWORDS  calmodulin binding; chromoprotein; FAD; flavoprotein; FMN;
           heme; iron; NADP; oxidoreductase
SUMMARY    #length 230 #checksum 5579

Query Match 73.5%; Score 50; DB 2; Length 230;
Best Local Similarity 60.0%; Pred. NO. 2.35e+00;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 217 LPPGVTRQAL 226
   :|:|:|:|
QY 1 LPPGSTKRAL 10

RESULT 12
ENTRY  A39344 #type complete
TITLE  tumor-associated mucin (MUC1) homolog precursor - mouse
ORGANISM #formal_name Mus musculus #common_name house mouse

```

```

DATE       03-Apr-1992 #sequence_revision 03-Apr-1992 #text_change
ACCESSIONS A39344
REFERENCE  23-Feb-1997
#authors   Spicer, A.P.; Parry, G.; Patton, S.; Gendler, S.J.
#journal   J. Biol. Chem. (1991) 266:15099-15109
#title     Molecular cloning and analysis of the mouse homologue of the
           tumor-associated mucin, MUC1, reveals conservation of the
           potential O-glycosylation sites, transmembrane, and
           cytoplasmic domains and a loss of minisatellite-like
           polymorphism.
#cross-references MUID:91332029
#accession A39344
#status    preliminary
#molecule_type DNA
#residues  1-630 #label SPI
#cross-references GB:M64928
KEYWORDS  cytoskeleton; transmembrane protein
SUMMARY    #length 630 #molecular-weight 64622 #checksum 4588

Query Match 72.1%; Score 49; DB 2; Length 630;
Best Local Similarity 87.5%; Pred. NO. 3.89e+00;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 594 VPPGSTKR 601
   :|||||:
QY 1 LPPGSTKR 8

RESULT 13
ENTRY  S45167 #type complete
TITLE  chitin synthase (EC 2.4.1.16) 2 - yeast (Saccharomyces
           cerevisiae)
ALTERNATE_NAMES chitin-UDP acetyl-glucosaminyl-transferase 2; protein
           YBR038w; protein YBR0407
ORGANISM #formal_name Saccharomyces cerevisiae
DATE     17-May-1994 #sequence_revision 09-Sep-1994 #text_change
        04-Sep-1998
ACCESSIONS S45167; S45896; A30922
REFERENCE  Silverman, S.J.
           Yeast (1989) 5:459-467
           Similar and different domains of chitin synthases 1 and 2 of
           S. cerevisiae: two isozymes with distinct functions.
#cross-references MUID:90143137
#accession S45167
#molecule_type DNA
#residues  1-963 #label SIL
#cross-references EMBL:M23865; NID:g171219; PID:g171220
REFERENCE  S45893
#authors   Andre, B.; Cziepluch, C.; Hein, C.; Jauniaux, J.C.;
           Urrestarazu, A.; Vissers, S.
           A submission submitted to the Protein Sequence Database, August 1994
#accession S45896
#molecule_type DNA
#residues  1-963 #label AND
#cross-references EMBL:Z35907; NID:g536257; PID:g536258; MIPS:YBR038w
GENETICS   #gene SGD:CHS2
           #map_position 2R
           #cross-references SGD:S0000242; MIPS:YBR038w
FUNCTION    catalyzes the alpha-1,4-glycosylation of chitin by
           UDP-N-acetyl-D-glucosamine producing elongated chitin and
           UDP
CLASSIFICATION #superfamily chitin synthase csha
KEYWORDS  glycosyltransferase; hexosyltransferase; transmembrane
           protein
FEATURE     #domain transmembrane #status predicted #label TM1\
           #domain transmembrane #status predicted #label TM2\
           #domain transmembrane #status predicted #label TM3\
           #domain transmembrane #status predicted #label TM4\
           424-440
           644-660
           677-698
           708-732

```

```
||||:||||
QY 1 LPPGSTKRAL 10

RESULT 6
ENTRY JC6193 #type complete
TITLE tumor suppressor p53 - rabbit
ORGANISM #formal_name Oryctolagus cuniculus #common_name domestic
        rabbit
DATE 11-Apr-1997 #sequence_revision 09-May-1997 #text_change
        17-Mar-1999
ACCESSIONS JC6193
REFERENCE Le Goas, F.; May, P.; Ronco, P.; de Fromental, C.C.
#authors Gene (1997) 185:169-173
#journal cDNA cloning and immunological characterization of rabbit
#title p53.
#cross-references MUID:97208869
#accession JC6193
##molecule_type mRNA
##residues 1-391 ##label LEA
##cross-references EMBL:X90592; NID:g1532043; PID:e194962; PID:g1532044
GENETICS
#gene p53
CLASSIFICATION #superfamily cellular tumor antigen p53
KEYWORDS tumor
SUMMARY #length 391 #molecular-weight 43435 #checksum 4367

Query Match 95.68; Score 65; DB 2; Length 391;
Best Local Similarity 90.08; Pred. No. 6.10e-04;
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 296 LPPGSKRAL 305
||||:||||
QY 1 LPPGSTKRAL 10

RESULT 7
ENTRY JC6176 #type complete
TITLE tumor suppressor protein p53 - Chinese hamster
ORGANISM #formal_name Crictetus griseus #common_name Chinese hamster
DATE 11-Apr-1997 #sequence_revision 09-May-1997 #text_change
        08-Sep-1997
ACCESSIONS JC6176
REFERENCE Lee, H.; Larner, J.M.; Hamlin, J.L.
#authors Gene (1997) 184:177-183
#journal Cloning and characterization of Chinese hamster p53 cDNA.
#title TATA-binding protein, and affects DNA replication, transcription,
#cross-references MUID:97183659
#contents liver
#accession JC6176
##molecule_type mRNA
##residues 1-393 ##label LEE
##cross-references GB:U50395; NID:g1842229; PID:g1842230
COMMENT This protein is a multimer, it plays the central role in a complex
        DNA damage-sensing network. It binds to replication factor and
        TATA-binding protein, and affects DNA replication, transcription,
        and recombination by protein/protein interactions.
GENETICS
#gene p53
CLASSIFICATION #superfamily cellular tumor antigen p53
KEYWORDS liver; tumor
SUMMARY #length 393 #molecular-weight 43362 #checksum 4043

Query Match 79.4%; Score 54; DB 2; Length 393;
Best Local Similarity 80.08; Pred. No. 2.93e-01;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 299 LPPKSARAL 308
||||:||||
QY 1 LPPGSTKRAL 10

RESULT 8
ENTRY JH0633 #type complete
TITLE cellular tumor antigen p53 - golden hamster
ORGANISM #formal_name Mesocricetus auratus #common_name golden hamster
DATE 17-Aug-1992 #sequence_revision 17-Aug-1992 #text_change
        08-Sep-1997
ACCESSIONS JH0633
REFERENCE Legros, Y.; McIntyre, P.; Soussi, T.
#authors Gene (1992) 112:247-250
#journal The cDNA cloning and immunological characterization of
#title hamster p53.
#cross-references MUID:92210007
#accession JH0633
##molecule_type mRNA
##residues 1-396 ##label LEG
##cross-references GB:M75144; NID:g191414; PID:g191415
##experimental_source kidney, strain MP1
GENETICS
#gene p53
CLASSIFICATION #superfamily cellular tumor antigen p53
KEYWORDS apoptosis; Cell division control; DNA binding; homotetramer;
        nucleus; phosphoprotein; transcription regulation; tumor
        suppressor; zinc
FEATURE 179,182,241,245 #binding_site zinc (Cys, His, Cys) #status
        predicted\
        395 #binding_site phosphoryl-RNA (Ser) (covalent) #status
        predicted
SUMMARY #length 396 #molecular-weight 43631 #checksum 6617

Query Match 79.4%; Score 54; DB 2; Length 396;
Best Local Similarity 80.0%; Pred. No. 2.93e-01;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 302 LPPKSARAL 311
||||:||||
QY 1 LPPGSTKRAL 10

RESULT 9
ENTRY S70581 #type complete
TITLE dihydropyrimidinase - rat
ORGANISM #formal_name Rattus norvegicus #common_name Norway rat
DATE 14-Feb-1997 #sequence_revision 13-Mar-1997 #text_change
        10-Sep-1997
ACCESSIONS S70581
REFERENCE Matsuda, K.; Sakata, S.; Kaneko, M.; Hamajima, N.; Nonaka,
#authors M.; Sasaki, M.; Tamaki, N.
#journal Biochim. Biophys. Acta (1996) 1307:140-144
#title Molecular cloning and sequencing of a cDNA encoding
        dihydropyrimidinase from the rat liver.
#cross-references MUID:96283806
#accession S70581
##status preliminary
##molecule_type mRNA
##residues 1-519 ##label MAT
##cross-references EMBL:D83704; NID:g1378018; PID:d1010479; PID:g1378019
SUMMARY #length 519 #molecular-weight 56833 #checksum 6037

Query Match 77.9%; Score 53; DB 2; Length 519;
Best Local Similarity 70.0%; Pred. No. 4.98e-01;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 40 LPPGDTSRGL 49
||||:||||
QY 1 LPPGSTKRAL 10

RESULT 10
ENTRY 152257 #type complete
```



#title Primary structure of DNA complementary to murine oncoprotein p53 mRNA.  
#cross-references MUID:88221682  
#accession S06336  
#status not compared with conceptual translation  
#molecule\_type mRNA  
#residues 1-134, 'V', 136-390 #label CHU  
REFERENCE  
#authors A02684  
Zakut-Houri, R.; Oren, M.; Bilenz, B.; Lavie, V.; Hazum, S.; Givol, D.  
#journal Nature (1983) 306:594-597  
#title A single gene and a pseudogene for the cellular tumour antigen p53.  
#cross-references MUID:84068204  
#accession A02684  
#molecule\_type mRNA  
#residues 1-159, 'H', 161-167, 'G', 169-233, 'I', 235-390 #label ZAK  
#cross-references GB:X01237; GB:K01700; NID:G53575  
REFERENCE  
#authors S38822  
Arai, N.; Nomura, D.; Yokota, K.; Wolf, D.; Brill, E.; Shohat, O.; Rotter, V.  
#journal Mol. Cell. Biol. (1986) 6:3232-3239  
#title Immunologically distinct p53 molecules generated by alternative splicing.  
#cross-references MUID:87064640  
#accession S38822  
#status preliminary  
#molecule\_type mRNA  
#residues 1-390 #label ARA1  
#cross-references EMBL:M13872; NID:G200198; PID:G200199  
#accession S38823  
#status preliminary  
#molecule\_type mRNA  
#residues 1-167, 'G', 169-233, 'I', 235-390 #label ARA2  
#cross-references EMBL:M13873  
REFERENCE  
#authors S40014  
Arai, N.; Nomura, D.; Yokota, K.; Wolf, D.; Brill, E.; Shohat, O.; Rotter, V.  
#submission submitted to the EMBL Data Library, July 1988  
#accession S40014  
#molecule\_type mRNA  
#residues 1-167, 'G', 169-390 #label ARA3  
#cross-references EMBL:M13873; NID:G200200; PID:G200201  
REFERENCE  
#authors I48703  
Jenkins, J.R.; Rudge, K.; Redmond, S.; Wade-Evans, A.  
#journal Nucleic Acids Res. (1984) 12:5609-5626  
#title Cloning and expression analysis of full length mouse cDNA sequences encoding the transformation associated protein p53.  
#cross-references MUID:84272240  
#accession I48703  
#status preliminary; translated from GB/EMBL/DBJ  
#molecule\_type mRNA  
#residues 1-47, 'R', 49-78, 'QW', 82-390 #label RES  
#cross-references EMBL:X00743; NID:G53570; PID:G53571  
COMMENT This DNA-binding protein plays an essential role in the regulation of cell division, as it is required for the transition from phase G0 to G1 of the cell cycle.  
COMMENT The tetramer association region may exhibit a beta-turn, beta-sheet, beta-turn, alpha-helix motif.  
CLASSIFICATION #superfamily cellular tumor antigen p53  
KEYWORDS apoptosis; cell division control; DNA binding; homotetramer; phosphoprotein; transcription regulation; tumor suppressor; zinc  
FEATURE  
1-44 #domain transcription activation #status predicted  
#label TRA  
16-26 #region conserved region I\  
99-289 #domain DNA-binding core #status predicted #label DBC\  
108-121 #region L1 loop\  
114-139 #region conserved region II\  
160-192 #region L2 loop\  
168-178 #region conserved region III\  
#region conserved region IV\  
#region L3 loop\  
#region conserved region V\  
#region nuclear location signal\  
#region tetramer association\  
#binding\_site phosphate (Ser) (covalent) #status predicted\  
#binding\_site zinc (Cys, His, Cys, Cys) #status predicted\  
#binding\_site phosphate (Ser) (covalent) (by cdcd kinase) #status predicted\  
#binding\_site phosphoryl-RNA (Ser) (covalent) #status predicted  
SUMMARY #length 390 #molecular-weight 43458 #checksum 1260  
Query Match 95.6%; Score 65; DB 1; Length 390;  
Best Local Similarity 90.0%; Pred. No. 6.10e-04;  
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 296 LPPGSAKRAL 305  
||||:||||  
QY 1 LPPGSTRKAL 10  
RESULT 5  
ENTRY S02192 #type complete  
TITLE cellular tumor antigen p53 - rat  
ALTERNATE\_NAMES gene p53 protein; nuclear oncoprotein p53  
ORGANISM #formal\_name Rattus norvegicus #common\_name Norway rat  
DATE 18-Oct-1989 #sequence\_revision 18-Oct-1989 #text\_change 17-Mar-1999  
ACCESSIONS S02192; S41149  
REFERENCE S02192  
#authors Soussi, T.; de Fromental, C.C.; Breugnot, C.; May, E.  
#journal Nucleic Acids Res. (1988) 16:11384  
#title Nucleotide sequence of a cDNA encoding the rat p53 nuclear oncoprotein.  
#cross-references MUID:89083585  
#accession S02192  
#molecule\_type mRNA  
#residues 1-391 #label SOU  
#cross-references EMBL:X13058; NID:G56828; PID:G56829  
REFERENCE S41149  
#authors Hulla, J.E.; Schneider, R.P.  
#journal Nucleic Acids Res. (1993) 21:713-717  
#title Structure of the rat p53 tumor suppressor gene.  
#cross-references MUID:93181268  
#accession S41149  
#status preliminary; nucleic acid sequence not shown; translation not shown  
#molecule\_type DNA  
#residues 1-173, 'W', 175-391 #label HUL  
#cross-references EMBL:L07909  
#note the nucleotide sequence was submitted to the EMBL Data Library, December 1992  
GENETICS 25/2; 32/3; 123/3; 185/1; 259/2; 305/1; 329/3; 365/2  
#introns #superfamily cellular tumor antigen p53  
CLASSIFICATION apoptosis; cell division control; DNA binding; homotetramer; nucleus; phosphoprotein; transcription regulation; tumor suppressor; zinc  
KEYWORDS  
FEATURE  
174,177,236,240 #binding\_site zinc (Cys, His, Cys, Cys) #status predicted\  
390 #binding\_site phosphoryl-RNA (Ser) (covalent) #status predicted  
SUMMARY #length 391 #molecular-weight 43451 #checksum 7105  
Query Match 95.6%; Score 65; DB 2; Length 391;  
Best Local Similarity 90.0%; Pred. No. 6.10e-04;  
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 297 LPPGSAKRAL 306

```
#authors Yamada, Y.; Yoshida, T.; Hayashi, K.; Sekiya, T.; Yokota, J.;
Hirohashi, S.; Nakatani, K.; Nakano, H.; Sugimura, T.;
Terada, M.
#journal Cancer Res. (1991) 51:5800-5805
#title p53 gene mutations in gastric cancer metastases and in
gastric cancer cell lines derived from metastases.
#cross-references MUID:92034678
#accession A44905
...
Note: remainder of annotations omitted.
Query Match 100.0%; Score 68; DB 1; Length 393;
Best Local Similarity 100.0%; Pred. No. 1.03e-04;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 299 LPPGSTKRAL 308
|||||
Qy 1 LPPGSTKRAL 10

RESULT 2
ENTRY S06594 #type complete
TITLE cellular tumor antigen p53 - green monkey
ORGANISM #formal_name Cercopithecus aethiops #common_name green
monkey, grivet
DATE 28-Feb-1990 #sequence_revision 28-Feb-1990 #text_change
ACCESSIONS S06594
REFERENCE S06594
#authors Rigaudy, P.; Eckhart, W.
#journal Nucleic Acids Res. (1989) 17:8375
#title Nucleotide sequence of a cDNA encoding the monkey cellular
phosphoprotein p53.
#cross-references MUID:90045967
#accession S06594
#molecule_type mRNA
#residues 1-393 #label RIG
#cross-references EMBL:X16384; NID:g22795; PID:g22796
CLASSIFICATION #superfamily cellular tumor antigen p53
KEYWORDS apoptosis; cell division control; DNA binding; homotetramer;
nucleus; phosphoprotein; transcription regulation; tumor
suppressor; zinc
FEATURE
176,179,238,242 #binding_site zinc (Cys, His, Cys, Cys) #status
predicted\
392 #binding_site phosphoryl-RNA (Ser) (covalent) #status
Predicted\
SUMMARY #length 393 #molecular-weight 43696 #checksum 4263
Query Match 100.0%; Score 68; DB 2; Length 393;
Best Local Similarity 100.0%; Pred. No. 1.03e-04;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 299 LPPGSTKRAL 308
|||||
Qy 1 LPPGSTKRAL 10

RESULT 3
ENTRY S38824 #type complete
TITLE cellular tumor antigen p53, minor splice form - mouse
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change
17-Mar-1999
ACCESSIONS S38824; S35478
REFERENCE S38822
#authors Arai, N.; Nomura, D.; Yokota, K.; Wolf, D.; Brill, E.;
Shohat, O.; Rotter, V.
#journal Mol. Cell. Biol. (1986) 6:3232-3239
#title Immunologically distinct p53 molecules generated by
alternative splicing.
#cross-references MUID:87064640
#accession S38824
```

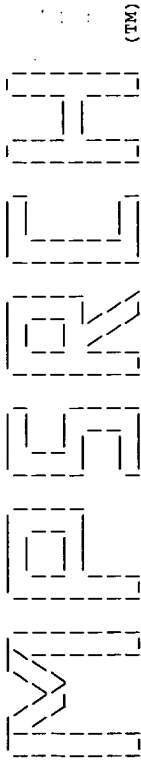
```
#molecule_type mRNA
#residues 1-381 #label ARA
#cross-references GB:M13874; NID:g200202; PID:g200203
REFERENCE S35478
#authors Han, K.A.; Kulesz-Martin, M.F.
#journal Nucleic Acids Res. (1992) 20:1979-1981
#title Alternatively spliced p53 RNA in transformed and normal cells
Of different tissue types.
#cross-references MUID:92253421
#accession S35478
#status nucleic acid sequence not shown; translation not shown
#molecule_type mRNA
#residues 1-381 #label HAN
#cross-references EMBL:M13874; NID:g200202; PID:g200203
#note the nucleotide sequence was submitted to the EMBL Data
Library, July 1988
COMMENT This sequence, produced by alternative splicing of the tenth
intron, lacks the carboxyl-terminal sequence necessary for
covalent attachment of RNA. The function of this minor splice
form is not known.
CLASSIFICATION #superfamily cellular tumor antigen p53
KEYWORDS alternative splicing; phosphoprotein; zinc
FEATURE
1-44 #domain transcription activation #status predicted
#label TRA\
16-26 #region conserved region I\
99-289 #domain DNA-binding core #status predicted #label DBC\
108-121 #region L1 loop\
114-139 #region conserved region II\
160-192 #region L2 loop\
168-178 #region conserved region III\
231-252 #region conserved region IV\
233-248 #region L3 loop\
267-283 #region conserved region V\
313-319 #region nuclear location signal\
319-357 #region tetramer association\
7,9,12,18,23,37 #binding_site phosphate (Ser) (covalent) #status
predicted\
173,176,235,239 #binding_site zinc (Cys, His, Cys, Cys) #status
predicted\
312 #binding_site phosphate (Ser) (covalent) (by cdcd2
kinase) #status predicted
SUMMARY #length 381 #molecular-weight 42498 #checksum 8703
Query Match 95.6%; Score 65; DB 2; Length 381;
Best Local Similarity 90.0%; Pred. No. 6.10e-04;
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Db 296 LPPGSTKRAL 305
|||||
Qy 1 LPPGSTKRAL 10

RESULT 4
ENTRY DNMS53 #type complete
TITLE cellular tumor antigen p53 - mouse
ALTERNATE_NAMES oncoprotein p53
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE 28-Aug-1985 #sequence_revision 04-Oct-1996 #text_change
12-Feb-1999
ACCESSIONS A22739; S06336; A02684; S38822; S38823; S40014; I48703
REFERENCE A22739
#authors Blenz, B.; Zakut-Houri, R.; Givol, D.; Oren, M.
#journal EMBO J. (1984) 3:2179-2183
#cross-references MUID:85027173
#accession A22739
#molecule_type DNA
#residues 1-134, 'V', 136-390 #label BIE
#cross-references GB:X00876; NID:g871420; PID:g871421; GB:X01237;
GB:K01700; NID:g53575; PID:g53576
REFERENCE S06336
#authors Chumakov, P.M.
#journal Bioorg. Khim. (1987) 13:1691-1694
```

```
##cross-references EMBL:X54156; NID:g35213; PID:g35214
REFERENCE S42669
#authors Matlashewski, G.; Lamb, P.; Pim, D.; Peacock, J.; Crawford, L.; Benchimol, S.
#journal EMBO J. (1994) 3:3257-3262
#title Isolation and characterization of a human p53 cDNA clone: expression of the human p53 gene.
#cross-references MUID:85126934
#accession S42669
##molecule_type mRNA
##residues 101-393 #label MK11
##cross-references EMBL:X01405; NID:g35215; PID:g642241
REFERENCE A22837
#authors Zakut-Houri, R.; Bienz-Tadmor, B.; Givol, D.; Oren, M.
#journal EMBO J. (1985) 4:1251-1255
#title Human p53 cellular tumor antigen: cDNA sequence and expression in COS cells.
#cross-references MUID:85230577
#accession A22837
##molecule_type mRNA
##residues 1-71,'P',73-393 #label ZAK
##cross-references EMBL:X02469; EMBL:M60950; NID:g35210
REFERENCE A55060
#authors Harlow, E.; Williamson, N.M.; Ralston, R.; Helfman, D.M.; Adams, T.E.
#journal Mol. Cell. Biol. (1985) 5:1601-1610
#title Molecular cloning and in vitro expression of a cDNA clone for human cellular tumor antigen p53.
#cross-references MUID:85267676
#accession A55060
##molecule_type mRNA
##residues 1-71,'P',73-272,'H',274-393 #label HAR
##cross-references GB:X03199; NID:gi89478; PID:gi89479
##experimental_source clone pr4-2, cell line A431
REFERENCE A93086
#authors Harris, N.; Brill, E.; Shohat, O.; Prokocimer, M.; Wolf, D.; Arai, N.; Rotter, V.
#journal Mol. Cell. Biol. (1986) 6:4650-4656
#title Molecular basis for heterogeneity of the human p53 protein.
#cross-references MUID:87089826
#accession A25397
##molecule_type mRNA
##residues 1-78,'P',80-393 #label HAR1
##cross-references EMBL:M14694; NID:g339813; PID:g339814
##experimental_source clone p53-H-1, transformed hybridoma SV-80 cell line
#accession B25397
##molecule_type mRNA
##residues 1-71,'P',73-78,'M',80-393 #label HAR2
##cross-references EMBL:M14695; NID:g339815; PID:g339816
##experimental_source clone p53-H-19, transformed hybridoma SV-80 cell line
REFERENCE S42452
#authors Matlashewski, G.J.; Tuck, S.; Pim, D.; Lamb, P.; Schneider, J.; Crawford, L.V.
#journal Mol. Cell. Biol. (1987) 7:961-963
#title Primary structure polymorphism at amino acid residue 72 of human p53.
#cross-references MUID:87144273
#accession S42452
##molecule_type mRNA; DNA
##residues 66-71,'P',73-79 #label MK12
##experimental_source clone lambda C113
##note 72-Cys was also found, and appears to represent a polymorphism
#accession S42453
##molecule_type mRNA; DNA
##residues 66-79 #label MK13
##experimental_source clone J6K
REFERENCE I38082
#authors Farrell, P.J.; Allan, G.J.; Shanahan, F.; Vousden, K.H.; Crook, T.
#journal EMBO J. (1991) 10:2879-2887
```

```
#title p53 is frequently mutated in Burkitt's lymphoma cell lines.
#cross-references MUID:92007731
#accession I38082
##status translated from GB/EMBL/DBDJ
##molecule_type mRNA
##residues 1-189,'LLSILSEKKEICVWSIWTETLFDIVWCPMSRLRLALT', 'VPSSTTTCTVTPAWAA' #label F01
##cross-references EMBL:X60010; NID:g506432; PID:g506433
##note deletion of a C nucleotide causes a frameshift at position 566
#accession I38083
##status translated from GB/EMBL/DBDJ
##molecule_type mRNA
##residues 1-192,'R',194-393 #label F02
##cross-references EMBL:X60011; NID:g506434; PID:g506435
#accession I38084
##status translated from GB/EMBL/DBDJ
##molecule_type mRNA
##residues 1-393 #label F03
##cross-references EMBL:X60012; NID:g506436; PID:g506437
#accession I38085
##status translated from GB/EMBL/DBDJ
##molecule_type mRNA
##residues 1-245,'T',247-393 #label F04
##cross-references EMBL:X60013; NID:g506438; PID:g506439
#accession I38086
##status translated from GB/EMBL/DBDJ
##molecule_type mRNA
##residues 1-236,'I',238-393 #label F05
##cross-references EMBL:X60014; NID:g506440; PID:g506441
#accession I38087
##status translated from GB/EMBL/DBDJ
##molecule_type mRNA
##residues 1-247,'Q',249-393 #label F06
##cross-references EMBL:X60015; NID:g506442; PID:g506443
#accession I38088
##status translated from GB/EMBL/DBDJ
##molecule_type mRNA
##residues 1-71,'P',73-237,'Y',239-393 #label F07
##cross-references EMBL:X60016; NID:g506444; PID:g506445
#accession I38089
##status translated from GB/EMBL/DBDJ
##molecule_type mRNA
##residues 1-247,'Q',249-393 #label F08
##cross-references EMBL:X60017; NID:g506446; PID:g506447
#accession I38090
##status translated from GB/EMBL/DBDJ
##molecule_type mRNA
##residues 1-71,'P',73-162,'H',164-393 #label F09
##cross-references EMBL:X60018; NID:g506448; PID:g506449
#accession I38091
##status translated from GB/EMBL/DBDJ
##molecule_type mRNA
##residues 1-212,'Q',214-393 #label F10
##cross-references EMBL:X60019; NID:g506450; PID:g506451
#accession I38092
##status translated from GB/EMBL/DBDJ
##molecule_type mRNA
##residues 1-253,'D',255-393 #label F11
##cross-references EMBL:X60020; NID:g506452; PID:g506453
##note all sequences submitted to the EMBL/GenBank/DBDJ databases June 1991
REFERENCE I38093
#authors Futreal, P.A.; Barrett, J.C.; Wiseman, R.W.
#journal Nucleic Acids Res. (1991) 19:6977
#title An Alu polymorphism intragenic to the TP53 gene.
#cross-references MUID:92107726
#accession I38093
##status translated from GB/EMBL/DBDJ
##molecule_type DNA
##residues 1-393 #label FUT
##cross-references EMBL:X54156; NID:g35213; PID:g35214
REFERENCE A44905
```

\*\*\*\*\*



(TM)

\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Sat Apr 15 00:50:39 2000; MasPar time 3.17 Seconds  
Tabular output not generated. 126.362 Million cell updates/sec

Title: >US-08-452-843-18  
Description: (1-10) from US08452843.pep  
Perfect Score: 68  
Sequence: 1 LPPGSTRAL 10

Scoring table: PAM 150  
Gap 15

Searched: 122810 seqs, 40068593 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: p1r62  
1:p1r1 2:p1r2 3:p1r3 4:p1r4

Statistics: Mean 22.573; Variance 26.106; scale 0.865

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	68	100.0	393	1 DNHU53	cellular tumor antige	1.03e-04
2	68	100.0	393	2 S06594	cellular tumor antige	1.03e-04
3	65	95.6	381	2 S38824	cellular tumor antige	6.10e-04
4	65	95.6	390	1 DNM553	cellular tumor antige	6.10e-04
5	65	95.6	391	2 S02192	cellular tumor antige	6.10e-04
6	65	95.6	391	2 JC6193	tumor suppressor p53	6.10e-04
7	54	79.4	393	2 JC6176	tumor suppressor prot	2.93e-01
8	54	79.4	396	2 JH0633	cellular tumor antige	2.93e-01
9	53	77.9	519	2 S70581	dihydropyrimidinase -	4.98e-01
10	51	75.0	631	2 I52257	epistatin - mouse	1.41e+00
11	50	73.5	230	2 I56979	nitric-oxide synthase	2.35e+00
12	49	72.1	630	2 A39344	tumor-associated muc	3.89e+00
13	49	72.1	963	2 S45167	chitin synthase (EC 2	3.89e+00
14	49	72.1	1199	2 A41939	G protein-coupled glu	3.89e+00
15	49	72.1	2185	2 S06200	acetyl-CoA carboxylas	3.89e+00
16	48	70.6	139	2 S78253	ribosomal protein l13	6.39e+00
17	48	70.6	386	2 S51648	cellular tumor antige	6.39e+00
18	48	70.6	455	2 B36916	site-specific recombi	6.39e+00
19	48	70.6	523	1 A41648	protein-tyrosine-phos	6.39e+00
20	48	70.6	559	2 I49444	SH3 binding protein -	6.39e+00
21	47	69.1	323	2 A40433	prephorylase pyrophosp	1.04e+01
22	47	69.1	758	2 S65169	hypothetical protein	1.04e+01
23	47	69.1	1299	2 T00261	hypothetical protein	1.04e+01

24	46	67.6	154	2	T01755	hypothetical protein	1.69e+01
25	46	67.6	370	2	B35255	chloromuconate cyclo	1.69e+01
26	46	67.6	544	2	S06602	modulo antigen - fru	1.69e+01
27	46	67.6	631	2	T00925	hypothetical protein	1.69e+01
28	46	67.6	2233	2	S63347	acetyl-CoA carboxylas	1.69e+01
29	45	66.2	137	2	T00921	hypothetical protein	2.72e+01
30	45	66.2	181	2	I52731	gene mvhlh1 protein -	2.72e+01
31	45	66.2	321	2	S31711	alternative transcrip	2.72e+01
32	45	66.2	368	2	T03580	probable transcrip	2.72e+01
33	45	66.2	470	2	T02319	hypothetical protein	2.72e+01
34	45	66.2	576	2	S65001	probable membrane pro	2.72e+01
35	45	66.2	649	2	T01882	hypothetical protein	2.72e+01
36	45	66.2	653	2	A49722	endoglin precursor -	2.72e+01
37	45	66.2	1139	2	S61918	protein kinase C (EC	2.72e+01
38	45	66.2	1526	2	JN0598	DNA topoisomerase (AT	2.72e+01
39	45	66.2	3131	2	S39842	enolatin synthetase -	2.72e+01
40	44	64.7	96	2	D71075	hypothetical protein	4.33e+01
41	44	64.7	171	2	B70408	conserved hypothetical	4.33e+01
42	44	64.7	498	2	S45567	nuclear factor I-A -	4.33e+01
43	44	64.7	505	2	S01300	transcription factor,	4.33e+01
44	44	64.7	1502	1	RGBYH1	CYC1/CYP3 transcripti	4.33e+01
45	44	64.7	3020	2	A43932	mucin 2 precursor, in	4.33e+01

ALIGNMENTS

RESULT	1	DNHU53	#type complete
ENTRY		cellular tumor antigen p53 - human	
TITLE		cellular phosphoprotein p53; oncoprotein p53; transformation	
ALTERNATE_NAMES		suppressor p53; tumor suppressor p53	
ORGANISM		#formal_name Homo sapiens #common_name man	
DATE		05-Oct-1988 #sequence_revision 18-Nov-1994 #text_change	
		26-Feb-1999	
ACCESSIONS		A25224; A43073; JT0436; S40773; S42669; A22837; A55080;	
		A25397; B25397; S42452; S4453; I38082; I38083; I38084;	
		I38085; I38086; I38087; I38088; I38089; I38090; I38091;	
		I38092; I38093; A44905; I58354; I78850; I52681; S60153	
REFERENCE		A25224	
#authors		Lamb, P.; Crawford, L.	
#journal		Mol. Cell. Biol. (1986) 6:1379-1385	
#title		Characterization of the human p53 gene.	
#cross-references		EMBL:87064416	
#accession		A25224	
#molecule_type		DNA	
#residues		1-393	#label LAM
#cross-references		EMBL:X01405; GB:M13121; GB:N00032; NID:g189460;	
		PID:g386994	
REFERENCE		JT0436	
#authors		Buchman, V.L.; Chumakov, P.M.; Ninkina, N.N.; Samarina, O.P.;	
		Georgiev, G.P.	
#journal		Gene (1988) 70:245-252	
#title		A variation in the structure of the protein-coding region of	
		the human p53 gene.	
#cross-references		EMBL:89108008	
#accession		A43073	
#molecule_type		DNA	
#residues		1-393	#label BUC1
#cross-references		EMBL:M22898; NID:g189474	
#note		this 72-Arg allele appears to be about 5 times more	
		frequent than the 72-Pro allele	
#accession		JT0436	
#molecule_type		DNA	
#residues		1-71	#label BUC2
#cross-references		EMBL:M22898; NID:g189474; PID:g189476	
#note		this 72-Pro allele was found in both normal and	
		malignant cell lines	
REFERENCE		S40773	
#authors		Chumakov, P.M.; Almazov, V.P.; Jenkins, J.R.	
#submission		submitted to the EMBL Data Library, August 1990	
#accession		S40773	
#molecule_type		DNA	
#residues		1-393	#label CHU

PS Example 1: 59-61; 82pp; English.  
CC Modified p53 variant p53C273del1364-393 (W13976) has the tumour-  
CC derived cysteine 273 mutation (see also W13952) and a deletion  
CC of the C-terminal 30 amino acids of wild-type p53 (see also  
CC W13948). Cys273 is a Class I p53 tumour mutation that affects DNA  
CC binding. The C-terminal deletion, introduced by site-directed  
CC mutagenesis of p53 DNA, activates the DNA binding of the p53  
CC tumour mutant. This provides the means for pharmacological rescue  
CC of p53 function in cancer patients. Other modified p53 constructs  
CC (W13949-50, W13953-54, W13968-77) have also been produced. Nucleic  
CC acids coding for modified p53 can be used for cancer gene therapy.  
SQ Sequence 363 AA;

Query Match 100.0%; Score 68; DB 1; Length 363;  
Best Local Similarity 100.0%; Pred. No. 3.53e-01;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 299 LPPGSTKRAL 308  
Qy 1 LPPGSTKRAL 10  
|||||

RESULT 15  
ID W13971 standard; Protein; 363 AA.  
AC W13971.  
DT 25-JUN-1997 (first entry)  
DE Modified p53 variant p53R284del1364-393.  
KW p53; tumour suppressor; cancer; therapy; cell proliferation;  
KW apoptosis; protein engineering; DNA binding.  
OS Synthetic.  
PN W09710843-A1.  
PD 27-MAR-1997.  
PF 20-SEP-1996; U15188.  
PR 22-SEP-1995; US-004802.  
PR 21-AUG-1996; US-697221.  
PA (WIST-) WISTAR INST ANATOMY & BIOLOGY.  
PI Halazonetis TD;  
DR WPI; 97-202618/18.  
PT R284K modified p53 protein having DNA binding ability - useful in  
PT treatment of cancer  
PS Example 1: 51-52; 82pp; English.  
CC Modified p53 variant p53R284del1364-393 (W13971) has a Thr284 to Arg  
CC substn. (see also W13949) and a deletion of the C-terminal 30  
CC amino acids. The R284R substitution, introduced by site-directed  
CC mutagenesis of p53 DNA, provides a novel p53-DNA contact between a  
CC phosphate of the DNA backbone and p53. The C-terminal deletion  
CC permits in vitro DNA binding. The variant provides the means for  
CC pharmacological rescue of p53 function in cancer patients. Other  
CC modified p53 constructs (W13949-50, W13953-54, W13968-77) have also  
CC been produced. Nucleic acids coding for modified p53 can be used  
CC for cancer gene therapy.  
SQ Sequence 363 AA;

Query Match 100.0%; Score 68; DB 1; Length 363;  
Best Local Similarity 100.0%; Pred. No. 3.53e-01;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 299 LPPGSTKRAL 308  
Qy 1 LPPGSTKRAL 10  
|||||

Search completed: Sat Apr 15 00:50:22 2000  
Job time : 36 secs.

RESULT 11  
ID W13975 standard; Protein; 363 AA.  
AC W13975;  
DT 25-JUN-1997 (first entry)  
DE Modified p53 variant p53H273R284del364-393.  
KW p53; tumour suppressor; cancer; therapy; cell proliferation;  
KW apoptosis; protein engineering; DNA binding.  
OS Synthetic.  
PN WO9710843-A1.  
PD 27-MAR-1997.  
PF 20-SEP-1996; U15188.  
PR 22-SEP-1995; US-004802.  
PR 21-AUG-1996; US-697221.  
PA (WIST-) WISTAR INST ANATOMY & BIOLOGY.  
PI Halazonetis TD;  
DR WPI; 97-202618/18.  
PT R284K modified p53 protein having DNA binding ability - useful in  
PT treatment of cancer  
PS Example 1; 58-59; 82pp; English.  
CC Modified p53 variant p53H273R284del364-393 (W13975) has the tumour-  
CC derived His273 mutation (see also W13952), a Thr284 to Arg substn.  
CC (see also W13949) and a deletion of the 30 C-terminal amino acids  
CC of wild-type p53 (W13948). His273 is a Class I p53 tumour mutation  
CC that affects DNA binding. The T284R substitution, introduced by  
CC site-directed mutagenesis of p53-DNA, provides a novel p53-DNA  
CC contact between a phosphate of the DNA backbone and p53, and  
CC restores DNA binding. The C-terminal deletion permits in vitro  
CC RNA binding. The construct provides the means for pharmacological  
CC rescue of p53 function in cancer patients. Other modified p53  
CC constructs (W13949-50, W13953-54, W13968-77) have also been  
CC produced. Nucleic acids coding for modified p53 can be used for  
CC cancer gene therapy.  
SQ Sequence 363 AA;

Query Match 100.0%; Score 68; DB 1; Length 363;  
Best Local Similarity 100.0%; Pred. No. 3.53e-01;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 299 LPPGSTKRAL 308  
QY 1 LPPGSTKRAL 10  
|||||

RESULT 12  
ID W13973 standard; Protein; 363 AA.  
AC W13973;  
DT 25-JUN-1997 (first entry)  
DE Modified p53 variant p53Q248R284del364-393.  
KW p53; tumour suppressor; cancer; therapy; cell proliferation;  
KW apoptosis; protein engineering; DNA binding.  
OS Synthetic.  
PN WO9710843-A1.  
PD 27-MAR-1997.  
PF 20-SEP-1996; U15188.  
PR 22-SEP-1995; US-004802.  
PR 21-AUG-1996; US-697221.  
PA (WIST-) WISTAR INST ANATOMY & BIOLOGY.  
PI Halazonetis TD;  
DR WPI; 97-202618/18.  
PT R284K modified p53 protein having DNA binding ability - useful in  
PT treatment of cancer  
PS Example 1; 54-56; 82pp; English.  
CC Modified p53 variant p53Q248R284del364-393 (W13973) has the tumour-  
CC derived Gln248 mutation (see also W13951), a Thr284 to Arg substn.  
CC (see also W13949) and a deletion of the 30 C-terminal amino acids  
CC of wild-type p53 (W13948). Gln248 is a Class I p53 tumour mutation  
CC that affects DNA binding. The T284R substitution, introduced by  
CC site-directed mutagenesis of p53-DNA, provides a novel p53-DNA  
CC contact between a phosphate of the DNA backbone and p53, and  
CC restores DNA binding. The C-terminal deletion permits in vitro  
CC RNA binding. The construct provides the means for pharmacological

CC rescue of p53 function in cancer patients. Other modified p53  
CC constructs (W13949-50, W13953-54, W13968-77) have also been  
CC produced. Nucleic acids coding for modified p53 can be used for  
CC cancer gene therapy.  
SQ Sequence 363 AA;

Query Match 100.0%; Score 68; DB 1; Length 363;  
Best Local Similarity 100.0%; Pred. No. 3.53e-01;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 299 LPPGSTKRAL 308  
QY 1 LPPGSTKRAL 10  
|||||

RESULT 13  
ID W13974 standard; Protein; 363 AA.  
AC W13974;  
DT 25-JUN-1997 (first entry)  
DE Modified p53 variant p53H273del364-393.  
KW p53; tumour suppressor; cancer; therapy; cell proliferation;  
KW apoptosis; protein engineering; DNA binding.  
OS Synthetic.  
PN WO9710843-A1.  
PD 27-MAR-1997.  
PF 20-SEP-1996; U15188.  
PR 22-SEP-1995; US-004802.  
PR 21-AUG-1996; US-697221.  
PA (WIST-) WISTAR INST ANATOMY & BIOLOGY.  
PI Halazonetis TD;  
DR WPI; 97-202618/18.  
PT R284K modified p53 protein having DNA binding ability - useful in  
PT treatment of cancer  
PS Example 1; 56-57; 82pp; English.  
CC Modified p53 variant p53H273del364-393 (W13974) has the tumour-  
CC derived histidine 273 mutation (see also W13952) and a deletion  
CC of the C-terminal 30 amino acids of wild-type p53 (see also  
CC W13948). His273 is a Class I p53 tumour mutation that affects DNA  
CC binding. The C-terminal deletion, introduced by site-directed  
CC mutagenesis of p53-DNA, activates the DNA binding of the p53  
CC tumour mutant. This provides the means for pharmacological rescue  
CC of p53 function in cancer patients. Other modified p53 constructs  
CC (W13949-50, W13953-54, W13968-77) have also been produced. Nucleic  
CC acids coding for modified p53 can be used for cancer gene therapy.  
SQ Sequence 363 AA;

Query Match 100.0%; Score 68; DB 1; Length 363;  
Best Local Similarity 100.0%; Pred. No. 3.53e-01;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 299 LPPGSTKRAL 308  
QY 1 LPPGSTKRAL 10  
|||||

RESULT 14  
ID W13976 standard; Protein; 363 AA.  
AC W13976;  
DT 25-JUN-1997 (first entry)  
DE Modified p53 variant p53C273del364-393.  
KW p53; tumour suppressor; cancer; therapy; cell proliferation;  
KW apoptosis; protein engineering; DNA binding.  
OS Synthetic.  
PN WO9710843-A1.  
PD 27-MAR-1997.  
PF 20-SEP-1996; U15188.  
PR 22-SEP-1995; US-004802.  
PR 21-AUG-1996; US-697221.  
PA (WIST-) WISTAR INST ANATOMY & BIOLOGY.  
PI Halazonetis TD;  
DR WPI; 97-202618/18.  
PT R284K modified p53 protein having DNA binding ability - useful in  
PT treatment of cancer

CC is not, i.e. they are not inactivated by dominant negative or oncogenic  
CC mutants, nor by other cellular proteins (because the leucine zipper  
CC domain prevents formation of inactive mixed oligomers).  
SQ Sequence 363 AA;

Query Match 100.0%; Score 68; DB 1; Length 363;  
Best Local Similarity 100.0%; Pred. No. 3.53e-01;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 307 LPPGSTKRAL 316  
|||||  
QY 1 LPPGSTKRAL 10

## RESULT 8

ID W28480 standard; Protein; 363 AA.  
AC W28480;  
DT 25-NOV-1997 (first entry)  
DE Human p53 protein variant V-325H.  
KW Leucine zipper domain; L2B; oligomerisation domain; mutant; mutagen;  
KW substitution; replacement; transactivation; viral protein VP16; HSV;  
KW anti-oncogene; hyperproliferation; cancer; restenosis;  
KW tumour suppression; apoptosis.  
OS Chimeric - Homo sapiens.  
OS Chimeric - Herpes simplex virus.  
OS Synthetic.

FH Key Location/Qualifiers

FT misc\_difference 189  
FT /note= "Arg residue at position 182 of wild-type  
FT p53 has been mutated to His"

PN WO9704092-A1.

PD 06-FEB-1997.

PF 17-JUL-1996; F01111.

PR 19-JUL-1995; FR-008729.

PA (RHON ) RHONE POULENC RORER SA.

PI Bracco L, Conseiller E;

DR WPI; 97-132633/12.

PT New p53 variants e.g. with oligomerisation domain replaced by

PT leucine zipper - useful for treating hyper-proliferative disorders,

PT esp. cancer and restenosis

PS Claim 30; Page -; 133pp; French.

CC Claimed variants of protein p53 have at least part of the  
CC oligomerisation domain deleted and replaced by a leucine zipper  
CC domain. The mutants preferably also have at least part of the p53  
CC transactivation domain (amino acids 1-74) deleted and replaced by  
CC the transactivating domain (TD) from herpes simplex virus viral  
CC protein VP16 (amino acids 411-490). The present sequence is that of  
CC a specifically claimed p53 variant designated V-325H and comprising  
CC the VP16 TD, amino acids 75-325 of human wild-type p53 (but with  
CC Arg182 replaced by His) and a leucine zipper domain at the C-terminal.  
CC The p53 variants are more active and more stable tumour suppressors  
CC and apoptosis-inducing agents than wild-type p53 and are active where  
CC the wild-type protein is not, i.e. they are not inactivated by dominant  
CC negative or oncogenic mutants, nor by other cellular proteins (because  
CC the leucine zipper domain prevents formation of inactive mixed  
CC oligomers).

CC (Note: this sequence does not appear in the specification and has  
CC been produced by modifying the given sequence of variant V-325).  
SQ Sequence 363 AA;

Query Match 100.0%; Score 68; DB 1; Length 363;  
Best Local Similarity 100.0%; Pred. No. 3.53e-01;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 307 LPPGSTKRAL 316  
|||||  
QY 1 LPPGSTKRAL 10

## RESULT 9

ID W13954 standard; Protein; 363 AA.

AC W13954.

DT 25-JUN-1997 (first entry)

DE Modified p53 variant (del364-393).  
KW p53; tumour suppressor; cancer; therapy; cell proliferation;  
KW apoptosis; protein engineering; DNA binding.  
OS Synthetic.  
PN WO9710843-A1.  
PD 27-MAR-1997.  
PF 20-SEP-1996; U15188.  
PR 22-SEP-1995; US-004802.  
PR 21-AUG-1996; US-697221.  
PA (WIST-) WISTAR INST ANATOMY & BIOLOGY.  
PI Halazonetis TD;  
DR WPI; 97-202618/18.  
DR R284K modified p53 protein having DNA binding ability - useful in  
PT treatment of cancer  
PS Example 1; 49-51; 82pp; English.  
CC A modified p53 variant (W13954) comprises wild-type p53 (see  
CC also W13948) having a deletion of the C-terminal 30 amino acids,  
CC and is obtd. by site-directed mutagenesis of p53 DNA. Deletion of  
CC the p53 C-terminal 30 amino acids activates the DNA binding of  
CC common Class I p53 mutants (see also W13951-52). Novel modified  
CC p53 variants (W13949-50, W13953-54, W13968-77), some contg.  
CC C-terminal deletions, provide the means for pharmacological rescue  
CC of p53 function in cancer patients. Nucleic acids coding for  
CC modified p53 can be used for cancer gene therapy.  
SQ Sequence 363 AA;

Query Match 100.0%; Score 68; DB 1; Length 363;  
Best Local Similarity 100.0%; Pred. No. 3.53e-01;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 299 LPPGSTKRAL 308  
|||||  
QY 1 LPPGSTKRAL 10

## RESULT 10

ID W13972 standard; Protein; 363 AA.  
AC W13972;  
DT 25-JUN-1997 (first entry)  
DE Modified p53 variant p53Q248del1364-393.  
KW p53; tumour suppressor; cancer; therapy; cell proliferation;  
KW apoptosis; protein engineering; DNA binding.  
OS Synthetic.  
PN WO9710843-A1.  
PD 27-MAR-1997.  
PF 20-SEP-1996; U15188.  
PR 22-SEP-1995; US-004802.  
PR 21-AUG-1996; US-697221.  
PA (WIST-) WISTAR INST ANATOMY & BIOLOGY.  
PI Halazonetis TD;  
DR WPI; 97-202618/18.  
DR R284K modified p53 protein having DNA binding ability - useful in  
PT treatment of cancer  
PS Example 1; 53-54; 82pp; English.  
CC Modified p53 variant p53Q248del1364-393 (W13972) has the tumour-  
CC derived glutamine 248 mutation (see also W13951) and a deletion  
CC of the C-terminal 30 amino acids of wild-type p53 (see also  
CC W13948). Gln248 is a Class I p53 tumour mutation that affects DNA  
CC binding. The C-terminal deletion, introduced by site-directed  
CC mutagenesis of p53 DNA, activates the DNA binding of the p53  
CC tumour mutant. This provides the means for pharmacological rescue  
CC of p53 function in cancer patients. Other modified p53 constructs  
CC (W13949-54, W13953-54, W13968-77) have also been produced. Nucleic  
CC acids coding for modified p53 can be used for cancer gene therapy.  
SQ Sequence 363 AA;

Query Match 100.0%; Score 68; DB 1; Length 363;  
Best Local Similarity 100.0%; Pred. No. 3.53e-01;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 299 LPPGSTKRAL 308  
|||||  
QY 1 LPPGSTKRAL 10

CC (Note: this sequence does not appear in the specification and has  
 CC been produced by modifying the given sequence of variant 393-325).  
 SQ Sequence 353 AA;

Query Match 100.0%; Score 68; DB 1; Length 353;  
 Best Local Similarity 100.0%; Pred. No. 3.53e-01;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 297 LPPGSTKRAL 306

QY 1 LPPGSTKRAL 10  
 |||||

# RESULT 5

ID W13960 standard; Protein; 359 AA.  
 AC W13960;  
 DT 25-JUN-1997 (first entry)  
 DE Chimeric p53 protein.  
 KW p53; tumour suppressor; cancer; therapy; cell proliferation;  
 KW apoptosis; protein engineering; GCN4; DNA binding.  
 OS Chimeric Homo sapiens;  
 OS Chimeric synthetic.  
 FH Key Location/Qualifiers  
 FT region 1..323  
 FT /label= p53wt  
 FT /note= "amino acids 1-323 of wild-type p53"  
 FT 324..326  
 FT /label= Linker  
 FT region 327..359  
 FT /label= GCN4  
 FT /note= "amino acids 249-281 of GCN4 LZ variant"

PN WO9710843-A1.  
 PD 27-MAR-1997.  
 PE 20-SEP-1996; U15188.  
 PR 22-SEP-1995; US-004802.  
 PR 21-AUG-1996; US-697221.  
 PA (WIST-) WISTAR INST ANATOMY & BIOLOGY.  
 PI Halazonetis TD;  
 DR WPI; 97-202618/18.  
 PT R284K modified p53 protein having DNA binding ability - useful in  
 PT treatment of cancer.  
 PS Disclosure; Refer to page 8; 82pp; English.  
 CC Chimeric p53 constructs (W13956-67) comprise N-terminal portions  
 CC of human wild-type p53 tumour suppressor (see also W13948) linked  
 CC to a C-terminal portion of the LZ variant (see also W13955) of  
 CC GCN4 and, in some cases, the C-terminal portion of wild-type  
 CC p53. The chimeric proteins have DNA binding activity and can  
 CC replace lost or insufficient p53 function, providing the means for  
 CC pharmacological rescue of p53 function in cancer patients. Nucleic  
 CC acids coding for modified p53 constructs can be used for cancer  
 CC gene therapy.  
 SQ Sequence 359 AA;

Query Match 100.0%; Score 68; DB 1; Length 359;  
 Best Local Similarity 100.0%; Pred. No. 3.53e-01;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 299 LPPGSTKRAL 308

QY 1 LPPGSTKRAL 10  
 |||||

# RESULT 6

ID W13961 standard; Protein; 361 AA.  
 AC W13961;  
 DT 25-JUN-1997 (first entry)  
 DE Chimeric p53 protein.  
 KW p53; tumour suppressor; cancer; therapy; cell proliferation;  
 KW apoptosis; protein engineering; GCN4; DNA binding.  
 OS Chimeric Homo sapiens;  
 OS Chimeric synthetic.  
 FH Key Location/Qualifiers  
 FT region 1..323

FT /label= p53wt  
 FT /note= "amino acids 1-323 of wild-type p53"  
 FT 324..329  
 FT /label= Linker  
 FT region 330..361  
 FT /label= GCN4  
 FT /note= "amino acids 250-281 of GCN4 LZ variant"  
 PN WO9710843-A1.  
 PD 27-MAR-1997.  
 PE 20-SEP-1996; U15188.  
 PR 22-SEP-1995; US-004802.  
 PR 21-AUG-1996; US-697221.  
 PA (WIST-) WISTAR INST ANATOMY & BIOLOGY.  
 PI Halazonetis TD;  
 DR WPI; 97-202618/18.  
 PT R284K modified p53 protein having DNA binding ability - useful in  
 PT treatment of cancer.  
 PS Disclosure; Refer to page 8; 82pp; English.  
 CC Chimeric p53 constructs (W13956-67) comprise N-terminal portions  
 CC of human wild-type p53 tumour suppressor (see also W13948) linked  
 CC to a C-terminal portion of the LZ variant (see also W13955) of  
 CC GCN4 and, in some cases, the C-terminal portion of wild-type  
 CC p53. The chimeric proteins have DNA binding activity and can  
 CC replace lost or insufficient p53 function, providing the means for  
 CC pharmacological rescue of p53 function in cancer patients. Nucleic  
 CC acids coding for modified p53 constructs can be used for cancer  
 CC gene therapy.  
 SQ Sequence 361 AA;

Query Match 100.0%; Score 68; DB 1; Length 361;  
 Best Local Similarity 100.0%; Pred. No. 3.53e-01;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 299 LPPGSTKRAL 308

QY 1 LPPGSTKRAL 10  
 |||||

# RESULT 7

ID W28479 standard; Protein; 363 AA.  
 AC W28479;  
 DT 25-NOV-1997 (first entry)  
 DE Human p53 protein variant V-325 encoded by pEC114.  
 KW Leucine zipper domain; LZD; oligomerisation domain; mutant; mutein;  
 KW substitution; replacement; transactivation; viral protein VP16; HSV;  
 KW anti-oncogene; hyperproliferation; cancer; restenosis;  
 KW tumour suppression; apoptosis.  
 OS Chimeric - Homo sapiens.  
 OS Chimeric - Herpes simplex virus.  
 OS Synthetic.  
 PN WO9704092-A1.  
 PD 06-FEB-1997.  
 PR 17-JUL-1996; F01111.  
 PR 19-JUL-1995; FR-008729.  
 PA (RHON) RHONE-POULENC RORER SA.  
 PI Bracco L, Conseiller E;  
 DR WPI; 97-132633/12.  
 DR N-PSDB; T86215.  
 PT New p53 variants e.g. with oligomerisation domain replaced by  
 PT leucine zipper - useful for treating hyper-proliferative disorders,  
 PT esp. cancer and restenosis  
 PS Claim 30; Pages 76-78; 133pp; French.  
 CC Claimed variants of protein p53 have at least part of the  
 CC oligomerisation domain deleted and replaced by a leucine zipper  
 CC domain. The mutants preferably also have at least part of the p53  
 CC transactivation domain (amino acids 1-74) deleted and replaced by  
 CC the transactivating domain (TD) from herpes simplex virus viral  
 CC protein VP16 (amino acids 411-490). The present sequence is that of  
 CC a specifically claimed p53 variant designated V-325 and comprising  
 CC the VP16 TD, amino acids 75-325 of human wild-type p53 and a  
 CC leucine zipper domain at the C-terminal. The p53 variants are  
 CC more active and more stable tumour suppressors and apoptosis-inducing  
 CC agents than wild-type p53 and are active where the wild-type protein



```

ID W28497 standard; Protein; 335 AA.
AC W28497;
DE 25-NOV-1997 (first entry)
KW Leucine zipper domain; LZD; oligomerisation domain; mutant; mutin;
KW substitution; replacement; transactivation; hinge region;
KW anti-oncogene; hyperproliferation; cancer; restenosis;
OS Homo sapiens.
FH Synthetic.
FT region Location/Qualifiers
FT 39..53
FT /label= hinge
FT
PN WO9704092-A1.
PD 06-FEB-1997.
PF 17-JUL-1996; F01111.
PR 19-JUL-1995; FR-008729.
PA (RHON ) RHONE POULENC RORER SA.
PI Bracco L, Conseiller E;
DR WPI; 97-132633/12.
DR N-PSDB; T86224.
PT New p53 variants e.g. with oligomerisation domain replaced by
PT leucine zipper - useful for treating hyper-proliferative disorders,
PT esp. cancer and restenosis
PS Claim 39; Pages 94-95; 133pp; French.
CC Claimed variants of protein p53 have at least part of the
CC oligomerisation domain deleted and replaced by a leucine zipper
CC domain. The mutants preferably also have at least part of the p53
CC transactivation domain (amino acids 1-74) deleted and replaced by
CC the domain 325-360 of p53. The present sequence is that of a
CC specifically claimed p53 variant designated 360h-325 and comprising
CC the 325-360 domain, separated from amino acids 75-325 of human
CC wild-type p53 (but with Arg182 replaced by His) by a synthetic hinge
CC sequence (Gly4Ser)3, and with a leucine zipper domain at the C-terminal.
CC The p53 variants are more active and more stable tumour suppressors
CC and apoptosis-inducing agents than wild-type p53 and are active where
CC the wild-type protein is not, i.e. they are not inactivated by dominant
CC negative or oncogenic mutants, nor by other cellular proteins (because
CC the leucine zipper domain prevents formation of inactive mixed
CC oligomers).
CC (Note: this sequence does not appear in the specification and has
CC been produced by modifying the given sequence of variant 360h-325).
SQ Sequence 335 AA;

Query Match 100.0%; Score 68; DB 1; Length 335;
Best Local Similarity 100.0%; Pred. No. 3.53e-01;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 279 LPPGSTRAL 288
QY 1 LPPGSTRAL 10
|||||
RESULT 4
ID W28494 standard; Protein; 353 AA.
AC W28494;
DE 25-NOV-1997 (first entry)
DE Human p53 protein variant 393-325H.
KW Leucine zipper domain; LZD; oligomerisation domain; mutant; mutin;
KW substitution; replacement; transactivation; viral protein VP16; HSV;
KW anti-oncogene; hyperproliferation; cancer; restenosis;
KW tumour suppression; apoptosis.
OS Homo sapiens.
FH Synthetic.
FT Key Location/Qualifiers
FT misc_difference 179
FT /note= "Arg residue at position 182 of wild-type
FT p53 has been mutated to His"
FT
PN WO9704092-A1.
PD 06-FEB-1997.
PF 17-JUL-1996; F01111.
PR 19-JUL-1995; FR-008729.
PA (RHON ) RHONE POULENC RORER SA.
PI Bracco L, Conseiller E;
DR WPI; 97-132633/12.
PT New p53 variants e.g. with oligomerisation domain replaced by
PT leucine zipper - useful for treating hyper-proliferative disorders,
PT esp. cancer and restenosis
PS Claim 37; Pages 94-95; 133pp; French.
CC Claimed variants of protein p53 have at least part of the
CC oligomerisation domain deleted and replaced by a leucine zipper
CC domain. The mutants preferably also have at least part of the p53
CC transactivation domain (amino acids 1-74) deleted and replaced by
CC the domain 325-393 of p53. The present sequence is that of a
CC specifically claimed p53 variant designated 393-325H and comprising
CC the 325-393 domain, amino acids 75-325 of human wild-type p53 (but with
CC Arg182 replaced by His) and a leucine zipper domain at the C-terminal.
CC The p53 variants are more active and more stable tumour suppressors
CC and apoptosis-inducing agents than wild-type p53 and are active where
CC the wild-type protein is not, i.e. they are not inactivated by dominant
CC negative or oncogenic mutants, nor by other cellular proteins (because
CC the leucine zipper domain prevents formation of inactive mixed
CC oligomers).

```

```

ID W28497 standard; Protein; 335 AA.
AC W28497;
DE 25-NOV-1997 (first entry)
KW Leucine zipper domain; LZD; oligomerisation domain; mutant; mutin;
KW substitution; replacement; transactivation; hinge region;
KW anti-oncogene; hyperproliferation; cancer; restenosis;
OS Homo sapiens.
FH Synthetic.
FT region Location/Qualifiers
FT 39..53
FT /label= hinge
FT
PN WO9704092-A1.
PD 06-FEB-1997.
PF 17-JUL-1996; F01111.
PR 19-JUL-1995; FR-008729.
PA (RHON ) RHONE POULENC RORER SA.
PI Bracco L, Conseiller E;
DR WPI; 97-132633/12.
DR N-PSDB; T86224.
PT New p53 variants e.g. with oligomerisation domain replaced by
PT leucine zipper - useful for treating hyper-proliferative disorders,
PT esp. cancer and restenosis
PS Claim 39; Pages 94-95; 133pp; French.
CC Claimed variants of protein p53 have at least part of the
CC oligomerisation domain deleted and replaced by a leucine zipper
CC domain. The mutants preferably also have at least part of the p53
CC transactivation domain (amino acids 1-74) deleted and replaced by
CC the domain 325-360 of p53. The present sequence is that of a
CC specifically claimed p53 variant designated 360h-325 and comprising
CC the 325-360 domain, separated from amino acids 75-325 of human
CC wild-type p53 by a synthetic hinge sequence (Gly4Ser)3, and with a
CC leucine zipper domain at the C-terminal. The p53 variants are
CC more active and more stable tumour suppressors and apoptosis-inducing
CC agents than wild-type p53 and are active where the wild-type protein
CC is not, i.e. they are not inactivated by dominant negative or oncogenic
CC mutants, nor by other cellular proteins (because the leucine zipper
CC domain prevents formation of inactive mixed oligomers).
SQ Sequence 335 AA;

Query Match 100.0%; Score 68; DB 1; Length 335;
Best Local Similarity 100.0%; Pred. No. 3.53e-01;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 279 LPPGSTRAL 288
QY 1 LPPGSTRAL 10
|||||
RESULT 3
ID W28498 standard; Protein; 335 AA.
AC W28498;
DE 25-NOV-1997 (first entry)
DE Human p53 protein variant 360h-325H.
KW Leucine zipper domain; LZD; oligomerisation domain; mutant; mutin;
KW substitution; replacement; transactivation; hinge region;
KW anti-oncogene; hyperproliferation; cancer; restenosis;
KW tumour suppression; apoptosis.
OS Homo sapiens.
FH Synthetic.
FT region Location/Qualifiers
FT 39..53
FT /label= hinge
FT
FT misc_difference 161
FT /note= "Arg residue at position 182 of wild-type
FT p53 has been mutated to His"
FT
PN WO9704092-A1.
PD 06-FEB-1997.
PF 17-JUL-1996; F01111.
PR 19-JUL-1995; FR-008729.
PA (RHON ) RHONE POULENC RORER SA.
PI Bracco L, Conseiller E;

```

\*\*\*\*\*

WATERMAN

(TM)

\*\*\*\*\*

Release 3.1A John F. Collins, BioComputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Sat Apr 15 00:49:46 2000; MasPar time 3.20 Seconds  
Tabular output not generated. 74.085 Million cell updates/sec

Title: >US-08-452-843-18  
Description: (1-10) from US08452843.pap  
Perfect Score: 68  
Sequence: 1 LPPGSTKRAL 10

Scoring table: PAM 150  
Gap 15

Searched: 188963 seqs, 23686106 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: a-geneseq36  
1:geneseqp

Statistics: Mean 16.666; Variance 45.734; scale 0.364

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	68	100.0	25	1 W14695	Human p53 regulatory d	3.53e-01
2	68	100.0	335	1 W28497	Human p53 protein vari	3.53e-01
3	68	100.0	335	1 W28498	Human p53 protein vari	3.53e-01
4	68	100.0	353	1 W28490	Human p53 protein vari	3.53e-01
5	68	100.0	359	1 W13960	Chimeric p53 protein.	3.53e-01
6	68	100.0	361	1 W13961	Chimeric p53 protein.	3.53e-01
7	68	100.0	363	1 W28479	Human p53 protein vari	3.53e-01
8	68	100.0	363	1 W28480	Human p53 protein vari	3.53e-01
9	68	100.0	363	1 W13974	Modified p53 variant (	3.53e-01
10	68	100.0	363	1 W13972	Modified p53 variant p	3.53e-01
11	68	100.0	363	1 W13975	Modified p53 variant p	3.53e-01
12	68	100.0	363	1 W13973	Modified p53 variant p	3.53e-01
13	68	100.0	363	1 W13974	Modified p53 variant p	3.53e-01
14	68	100.0	363	1 W13976	Modified p53 variant p	3.53e-01
15	68	100.0	363	1 W13971	Modified p53 variant p	3.53e-01
16	68	100.0	368	1 W13956	Chimeric p53 protein.	3.53e-01
17	68	100.0	374	1 W28482	Human p53 protein vari	3.53e-01
18	68	100.0	374	1 W28481	Human p53 protein vari	3.53e-01
19	68	100.0	381	1 W28489	Human p53 protein vari	3.53e-01
20	68	100.0	381	1 W28490	Human p53 protein vari	3.53e-01
21	68	100.0	393	1 Y03191	Amino acid sequence of	3.53e-01
22	68	100.0	393	1 W84270	Human p53 protein.	3.53e-01
23	68	100.0	393	1 W69218	Human p53 mutant 1.	3.53e-01

24	68	100.0	393	1 W69217	Human wild-type p53 pr	3.53e-01
25	68	100.0	393	1 W57244	Human p53 protein SEQ	3.53e-01
26	68	100.0	393	1 W05346	Human p53 mutant R273H	3.53e-01
27	68	100.0	393	1 W13968	Modified p53 variant p	3.53e-01
28	68	100.0	393	1 W05347	Human p53 mutant R248Q	3.53e-01
29	68	100.0	393	1 W13969	Modified p53 variant p	3.53e-01
30	68	100.0	393	1 W13970	Modified p53 variant p	3.53e-01
31	68	100.0	393	1 W25155	Human p53 variant foun	3.53e-01
32	68	100.0	393	1 W05349	Human p53 mutant R273C	3.53e-01
33	68	100.0	393	1 R91933	Wild type p53 protein.	3.53e-01
34	68	100.0	393	1 W02617	Human p53 tumour suppr	3.53e-01
35	68	100.0	393	1 W13978	Human tumour-derived p	3.53e-01
36	68	100.0	393	1 W13952	Human tumour-derived p	3.53e-01
37	68	100.0	393	1 W13951	Human tumour-derived p	3.53e-01
38	68	100.0	393	1 W13949	T284R modified human p	3.53e-01
39	68	100.0	401	1 W28487	Human p53 protein vari	3.53e-01
40	68	100.0	401	1 W28488	Human p53 protein vari	3.53e-01
41	68	100.0	404	1 W13963	Chimeric p53 protein.	3.53e-01
42	68	100.0	406	1 W13966	Chimeric p53 protein.	3.53e-01
43	68	100.0	406	1 W13964	Chimeric p53 protein.	3.53e-01
44	68	100.0	411	1 W13967	Chimeric p53 protein.	3.53e-01
45	68	100.0	533	1 W19763	p53-GM-CSF immunostimu	3.53e-01

ALIGNMENTS

RESULT 1  
ID W14695 standard; Peptide; 25 AA.  
AC W14695;1997 (first entry)  
DT 24-NOV-1997  
DE Human p53 regulatory domain I.  
KW Tumour suppressor protein; p53; cancer; hyperproliferation;  
KW therapy; mimetic; heat shock protein; Dnak.  
OS Homo sapiens.  
EH Key Location/Qualifiers  
FT binding\_site 4..13  
FT modified\_site 23  
FT /label= PAB421  
FT /label= Phosphorylation  
FT /note= "cdc2 phosphorylation site"  
PN W09714794-A1.  
PD 24-APR-1997.  
PF 21-OCT-1996; G02605.  
PR 20-OCT-1995; GB-021544.  
PA (UYDU-) UNIV DUNDEE.  
PI Hupp TR, Lane DP;  
DR WPI; 97-245111/22  
PT Substance which activates sequence specific DNA binding activity of  
PT latent p53 - useful for treatment of cancer or other  
PT hyperproliferative disorders  
PS Disclosure; Fig 15; 68pp; English.  
CC This peptide corresponds to amino acid residues 293-317 in the  
CC C-terminal negative regulatory domain of human tumour suppressor  
CC protein p53, and comprises regulatory domain I of p53. It is  
CC separated from regulatory domain II (W14697) by a tetramerisation  
CC domain. A binding site for monoclonal antibody PAB421, which  
CC activates p53 for DNA binding, is present in domain I. Regulatory  
CC domain II includes the Dnak binding site (see also W14694) of human  
CC p53. Substances that activate the DNA binding activity of latent  
CC p53 are useful in the treatment of cancer and other  
CC hyperproliferative disorders.  
SQ Sequence 25 AA;

Query Match 100.0%; Score 68; DB 1; Length 25;

Best Local Similarity 100.0%; Pred No. 3.53e-01;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 7 LPPGSTKRAL 16

Qy 1 LPPGSTKRAL 10

RESULT 2

**THIS PAGE BLANK (USPTO)**

```
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
DR EMBL; AF071574; AAD34216.1; -.
DR PROSITE; PS00348; P53; 1.
KW Anti-oncogene; DNA-binding; Transcription regulation; Activator;
KW Nuclear protein; Phosphorylation.
FT NON_TER 265 1
FT NON_TER 265 1
SQ SEQUENCE 265 AA; 29341 MW; 9C35CAC2 CRC32;

Query Match 100.0%; Score 64; DB 13; Length 265;
Best Local Similarity 100.0%; Pred. No. 3.41e-03;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 228 RPILTIITL 236
| | | | | | | |
QY 1 RPILTIITL 9

RESULT 13
ID Q29475 PRELIMINARY; PRT; 281 AA.
AC Q29475;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-NOV-1999 (TReMBLrel. 12, Last annotation update)
DE CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).
GN P53.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Carnivora; Fissipedia; Canidae; Canis.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-MAMMARY GLAND;
RX MEDLINE: 97194812.
RA VAN LEEUWEN I., RUTEMAN G.R., HELLMEN E., CORNELISSE C.C.J.,
RA DEVILLEE P.;
RT "P53 mutations in mammary tumor cell lines and corresponding tumor
RT tissues in the dog.";
RL Anticancer Res. 16:3737-3744(1996).
CC -1- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT
CC PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL
CC CYCLE REGULATION, IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY
CC REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED
CC FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF
CC CYCLIN-DEPENDENT KINASES (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
DR EMBL; L37107; AAC37335.1; -.
DR HSSP; P04637; ISAH.
DR PROSITE; PS00348; P53; 1.
DR PFAM; PF00870; P53; 1.
KW Anti-oncogene; DNA-binding; Transcription regulation; Activator;
KW Nuclear protein; Phosphorylation.
FT NON_TER 281 1
FT NON_TER 281 1
SQ SEQUENCE 281 AA; 31762 MW; FC7BAE31 CRC32;

Query Match 100.0%; Score 64; DB 6; Length 281;
Best Local Similarity 100.0%; Pred. No. 3.41e-03;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 142 RPILTIITL 150
| | | | | | | |
QY 1 RPILTIITL 9

RESULT 14
ID Q95326 PRELIMINARY; PRT; 285 AA.
AC Q95326;
DT 01-FEB-1997 (TReMBLrel. 02, Created)
DT 01-FEB-1997 (TReMBLrel. 02, Last sequence update)
DT 01-NOV-1999 (TReMBLrel. 12, Last annotation update)
DE CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).
GN P53.
OS Canis familiaris (Dog).
```

```
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Carnivora; Fissipedia; Canidae; Canis.
RN [1]
RP SEQUENCE FROM N.A.
RA YANG B.J., SHI X.B., LAU D.H.M.;
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT
CC PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL
CC CYCLE REGULATION, IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY
CC REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED
CC FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF
CC CYCLIN-DEPENDENT KINASES (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
DR EMBL; U62133; AAB16961.1; -.
DR HSSP; P04637; 1YCS.
DR PROSITE; PS00348; P53; 1.
DR PFAM; PF00870; P53; 1.
KW Anti-oncogene; DNA-binding; Transcription regulation; Activator;
KW Nuclear protein; Phosphorylation.
FT NON_TER 285 1
FT NON_TER 285 1
SQ SEQUENCE 285 AA; 31616 MW; 15E1EC47 CRC32;

Query Match 100.0%; Score 64; DB 6; Length 285;
Best Local Similarity 100.0%; Pred. No. 3.41e-03;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 204 RPILTIITL 212
| | | | | | | |
QY 1 RPILTIITL 9

RESULT 15
ID P50332 PRELIMINARY; PRT; 286 AA.
AC P50332;
DT 01-MAY-1997 (TReMBLrel. 03, Created)
DT 01-MAY-1997 (TReMBLrel. 03, Last sequence update)
DT 01-NOV-1999 (TReMBLrel. 12, Last annotation update)
DE P53 (FRAGMENT).
OS Mastomys natalensis papillomavirus (MnpV).
OC Viruses; dsDNA viruses, no RNA stage; Papovaviridae; Papillomavirus.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-SPONTANEOUS ECLOMAS;
RA LUQUE E.A., TANG L.H., MODLIN I.M.;
RL Gastroenterology 0:0-0(0).
DR EMBL; U48619; AAB41834.1; -.
DR HSSP; P04637; 1PET.
DR PFAM; PF00870; P53; 1.
FT NON_TER 286 1
FT NON_TER 286 1
SQ SEQUENCE 286 AA; 32247 MW; 5B5D3CAD CRC32;

Query Match 100.0%; Score 64; DB 14; Length 286;
Best Local Similarity 100.0%; Pred. No. 3.41e-03;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 142 RPILTIITL 150
| | | | | | | |
QY 1 RPILTIITL 9

Search completed: Sat Apr 15 00:46:41 2000
Job time : 95 secs.
```

```
Query Match      100.0%; Score 64; DB 11; Length 205;
Best Local Similarity 100.0%; Pred. No. 3.41e-03;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 124 RPILTIITL 132
|||||
QY 1 RPILTIITL 9

RESULT 9
ID P89004 PRELIMINARY; PRT; 238 AA.
AC P89004;
DT 01-MAY-1997 (TRENBLrel. 03, Created)
DT 01-MAY-1997 (TRENBLrel. 03, Last sequence update)
DT 01-NOV-1999 (TRENBLrel. 12, Last annotation update)
DE P53 (FRAGMENT).
OS Mastomys natalensis papillomavirus (MNPV).
OC Viruses; GSDNA viruses, no RNA stage; Papovaviridae; Papillomavirus.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-ECLONA INDUCED BY LOXTIDINE.;
RA LUQUE E.A., TANG L.H., MODLIN I.M.;
RL Gastroenterology 0:0-0(0).
DR EMBL; U48618; AAB41833.1; -.
DR HSSP; P04637; LYCS.
DR PFAM; PF00870; P53; 1.
FT NON_TER 1
SQ SEQUENCE 238 AA; 26704 MW; 097E01F9 CRC32;

Query Match      100.0%; Score 64; DB 14; Length 238;
Best Local Similarity 100.0%; Pred. No. 3.41e-03;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 142 RPILTIITL 150
|||||
QY 1 RPILTIITL 9

RESULT 10
ID Q9W681 PRELIMINARY; PRT; 265 AA.
AC Q9W681;
DT 01-NOV-1999 (TRENBLrel. 12, Created)
DT 01-NOV-1999 (TRENBLrel. 12, Last sequence update)
DT 01-NOV-1999 (TRENBLrel. 12, Last annotation update)
DE CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).
GN P53.
OS Oncorhynchus keta (Chum salmon).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Actinopterygii;
OC Neopterygii; Teleostei; Euteleostei; Protacanthopterygii;
OC Salmoniformes; Salmonidae; Oncorhynchus.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-OVARY;
RA BHASKARAN A., MAY D., RAND-WEAVER M., TYLER C.R.;
RT "Evolutionary conservancy of p53 gene sequences in fish.";
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT
CC PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL
CC CYCLE REGULATION, IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY
CC REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED
CC FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF
CC CYCLIN-DEPENDENT KINASES (BY SIMILARITY).
CC EMBL; AF071573; AAD34215.1; -.
DR PROSITE; PS00348; P53; 1.
DR Anti-oncogene; DNA-binding; Transcription regulation; Activator;
KW Nuclear protein; Phosphorylation.
FT NON_TER 1
FT NON_TER 265
SQ SEQUENCE 265 AA; 29376 MW; 16515773 CRC32;

Query Match      100.0%; Score 64; DB 13; Length 265;

Db 124 RPILTIITL 132
|||||
QY 1 RPILTIITL 9

RESULT 11
ID Q9W680 PRELIMINARY; PRT; 265 AA.
AC Q9W680;
DT 01-NOV-1999 (TRENBLrel. 12, Created)
DT 01-NOV-1999 (TRENBLrel. 12, Last sequence update)
DT 01-NOV-1999 (TRENBLrel. 12, Last annotation update)
DE CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).
GN P53.
OS Oncorhynchus kisutch (Coho salmon).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Actinopterygii;
OC Neopterygii; Teleostei; Euteleostei; Protacanthopterygii;
OC Salmoniformes; Salmonidae; Oncorhynchus.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-OVARY;
RA BHASKARAN A., MAY D., RAND-WEAVER M., TYLER C.R.;
RT "Evolutionary conservancy of p53 gene sequences in fish.";
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT
CC PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL
CC CYCLE REGULATION, IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY
CC REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED
CC FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF
CC CYCLIN-DEPENDENT KINASES (BY SIMILARITY).
CC EMBL; AF071572; AAD34214.1; -.
DR PROSITE; PS00348; P53; 1.
DR Anti-oncogene; DNA-binding; Transcription regulation; Activator;
KW Nuclear protein; Phosphorylation.
FT NON_TER 1
FT NON_TER 265
SQ SEQUENCE 265 AA; 29243 MW; 66872D86 CRC32;

Query Match      100.0%; Score 64; DB 13; Length 265;
Best Local Similarity 100.0%; Pred. No. 3.41e-03;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 228 RPILTIITL 236
|||||
QY 1 RPILTIITL 9

RESULT 12
ID Q9W682 PRELIMINARY; PRT; 265 AA.
AC Q9W682;
DT 01-NOV-1999 (TRENBLrel. 12, Created)
DT 01-NOV-1999 (TRENBLrel. 12, Last sequence update)
DT 01-NOV-1999 (TRENBLrel. 12, Last annotation update)
DE CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).
GN P53.
OS Oncorhynchus tshawytscha (Chinook salmon).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Actinopterygii;
OC Neopterygii; Teleostei; Euteleostei; Protacanthopterygii;
OC Salmoniformes; Salmonidae; Oncorhynchus.
RN [1]
RP SEQUENCE FROM N.A.
RA BHASKARAN A., MAY D., RAND-WEAVER M., TYLER C.R.;
RT "Evolutionary conservancy of p53 gene sequences in fish.";
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT
CC PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL
CC CYCLE REGULATION, IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY
CC REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED
CC FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF
CC CYCLIN-DEPENDENT KINASES (BY SIMILARITY).
CC EMBL; AF071573; AAD34215.1; -.
DR PROSITE; PS00348; P53; 1.
DR Anti-oncogene; DNA-binding; Transcription regulation; Activator;
KW Nuclear protein; Phosphorylation.
FT NON_TER 1
FT NON_TER 265
SQ SEQUENCE 265 AA; 29376 MW; 16515773 CRC32;
```

```
RA STWARSKI D., MAI S., SCHNEIDERMAN M.H., HUPPI K.;
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT
CC PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL
CC CYCLE REGULATION, IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY
CC REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED
CC FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF
CC CYCLIN-DEPENDENT KINASES (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
DR EMBL; U41451; BAB41266.1; -.
DR HSPSP; P04637; ITSR.
DR PROSITE; PS00348; P53; 1.
DR PFAM; PF00870; P53; 1.
KW Nuclear protein; Phosphorylation; Anti-oncogene; DNA-binding;
FT Transcription regulation; Activator.
FT NON_TER 1 1
FT NON_TER 136 136
SQ SEQUENCE 136 AA; 13411 MW; CFB916C9 CRC32;

Query Match 100.0%; Score 64; DB 11; Length 136;
Best Local Similarity 100.0%; Pred. No. 3.41e-03;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 118 RPILTIITL 126
   |||||
Qy 1 RPILTIITL 9

RESULT 6
ID Q29469 PRELIMINARY; PRT; 146 AA.
AC Q29469;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).
GN P53.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Carnivora; Fissipedia; Canidae; Canis.
RN [1]
RP SEQUENCE FROM N.A.
RA NASIR L., MCFARLANE S.T., ARGYLE D.J., REID S.W.J.;
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT
CC PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL
CC CYCLE REGULATION, IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY
CC REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED
CC FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF
CC CYCLIN-DEPENDENT KINASES (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
DR EMBL; U51857; AAD12203.1; -.
DR HSPSP; P04637; ITSR.
DR PROSITE; PS00348; P53; 1.
DR PFAM; PF00870; P53; 1.
KW Anti-oncogene; DNA-binding; Transcription regulation; Activator;
FT Nuclear protein; Phosphorylation.
FT NON_TER 1 1
FT NON_TER 146 146
SQ SEQUENCE 146 AA; 16396 MW; 8AE726C9 CRC32;

Query Match 100.0%; Score 64; DB 6; Length 146;
Best Local Similarity 100.0%; Pred. No. 3.41e-03;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 115 RPILTIITL 123
   |||||
Qy 1 RPILTIITL 9

RESULT 7
ID Q29484 PRELIMINARY; PRT; 196 AA.
AC Q29484;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
```

```
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DE 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
GN CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).
OS Equus caballus (Horse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Perissodactyla; Equidae; Equus.
RN [1]
RP SEQUENCE FROM N.A.
RA BUCHER K., SZALAI G., MARTI E., PAULI U., LAZARY S.;
RL Res. Vet. Sci. 0:0-0(0).
CC -1- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT
CC PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL
CC CYCLE REGULATION, IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY
CC REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED
CC FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF
CC CYCLIN-DEPENDENT KINASES (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
DR EMBL; X91793; CAAG2905.1; -.
DR HSPSP; P04637; ISAH.
DR PROSITE; PS00348; P53; 1.
DR PFAM; PF00870; P53; 1.
KW Anti-oncogene; DNA-binding; Transcription regulation; Activator;
KW Nuclear protein; Phosphorylation.
FT NON_TER 1 1
FT NON_TER 196 196
SQ SEQUENCE 196 AA; 22080 MW; F443239C CRC32;

Query Match 100.0%; Score 64; DB 6; Length 196;
Best Local Similarity 100.0%; Pred. No. 3.41e-03;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 116 RPILTIITL 124
   |||||
Qy 1 RPILTIITL 9

RESULT 8
ID Q35873 PRELIMINARY; PRT; 205 AA.
AC Q35873;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).
GN P53.
OS Cricetus griseus (Chinese hamster).
OC Eukaryota; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae; Cricetulus.
RN [1]
RP SEQUENCE FROM N.A.
RA RAINALDI G., MARCHETTI S., CAPECCHI B., MENEVERI R., PIRAS A.,
RA LEUZZI R.;
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA VATTERONI L., MUSIO A., MENEVERI R., RAINALDI G.;
RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT
CC PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL
CC CYCLE REGULATION, IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY
CC REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED
CC FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF
CC CYCLIN-DEPENDENT KINASES (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
DR EMBL; U74487; BAB82420.1; -.
DR HSPSP; P04637; ISAH.
DR PROSITE; PS00348; P53; 1.
DR PFAM; PF00870; P53; 1.
KW Anti-oncogene; DNA-binding; Transcription regulation; Activator;
KW Nuclear protein; Phosphorylation.
FT NON_TER 1 1
FT NON_TER 205 205
SQ SEQUENCE 205 AA; 23122 MW; 680DDDDC CRC32;
```

1998

\*\*\*\*\*

W P S R E L H (TM)

\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Sat Apr 15 00:45:06 2000; Maspar time 7.08 Seconds

Tabular output not generated. 88.104 Million cell updates/sec

Title: >US-08-452-843-17  
Description: (1-9) from US08452843.pep  
Perfect Score: 64  
Sequence: 1 RPILTIITL 9

Scoring table: PAM 150  
Gap 15

Searched: 225878 seqs, 69334122 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: sptrmb12  
1:sp\_archaea 2:sp\_bacteria 3:sp\_fungi 4:sp\_human  
5:sp\_invertebrate 6:sp\_mammal 7:sp\_mhc 8:sp\_Organelle  
9:sp\_phage 10:sp\_plant 11:sp\_rodent 12:sp\_unclassified  
13:sp\_vertebrate 14:sp\_virus

Statistics: Mean 23.640; Variance 26.527; scale 0.891

pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	64	100.0	37	11	Q64447 CELLULAR TUMOR ANTIGEN	3.41e-03
2	64	100.0	42	6	Q29446 CELLULAR TUMOR ANTIGEN	3.41e-03
3	64	100.0	135	11	Q64451 CELLULAR TUMOR ANTIGEN	3.41e-03
4	64	100.0	136	11	Q60434 CELLULAR TUMOR ANTIGEN	3.41e-03
5	64	100.0	136	11	Q64396 CELLULAR TUMOR ANTIGEN	3.41e-03
6	64	100.0	146	6	Q29459 CELLULAR TUMOR ANTIGEN	3.41e-03
7	64	100.0	196	6	Q29484 CELLULAR TUMOR ANTIGEN	3.41e-03
8	64	100.0	205	11	Q35873 CELLULAR TUMOR ANTIGEN	3.41e-03
9	64	100.0	238	14	P53 (FRAGMENT)	3.41e-03
10	64	100.0	265	13	Q9W681 CELLULAR TUMOR ANTIGEN	3.41e-03
11	64	100.0	265	13	Q9W680 CELLULAR TUMOR ANTIGEN	3.41e-03
12	64	100.0	265	13	Q9W682 CELLULAR TUMOR ANTIGEN	3.41e-03
13	64	100.0	281	6	Q29475 CELLULAR TUMOR ANTIGEN	3.41e-03
14	64	100.0	285	6	Q95326 CELLULAR TUMOR ANTIGEN	3.41e-03
15	64	100.0	286	14	P90332 P53 (FRAGMENT)	3.41e-03
16	64	100.0	286	14	P89003 P53 (FRAGMENT)	3.41e-03
17	64	100.0	376	13	Q93379 CELLULAR TUMOR ANTIGEN	3.41e-03
18	64	100.0	378	14	P89002 P53 (FRAGMENT)	3.41e-03
19	64	100.0	390	11	O70366 CELLULAR TUMOR ANTIGEN	3.41e-03
20	64	100.0	391	11	Q9WUR6 CELLULAR TUMOR ANTIGEN	3.41e-03

21	64	100.0	391	6	O36006 CELLULAR TUMOR ANTIGEN	3.41e-03
22	64	100.0	393	4	Q16535 P53 TRANSFORMATION SUP	3.41e-03
23	64	100.0	393	4	Q15087 P53 TRANSFORMATION SUP	3.41e-03
24	64	100.0	393	4	Q16808 CELLULAR TUMOR ANTIGEN	3.41e-03
25	64	100.0	393	4	Q16811 CELLULAR TUMOR ANTIGEN	3.41e-03
26	64	100.0	393	4	Q16848 CELLULAR TUMOR ANTIGEN	3.41e-03
27	64	100.0	393	4	Q16807 CELLULAR TUMOR ANTIGEN	3.41e-03
28	64	100.0	393	4	Q15088 P53 TRANSFORMATION SUP	3.41e-03
29	64	100.0	393	4	Q15086 P53 TRANSFORMATION SUP	3.41e-03
30	64	100.0	393	4	Q16809 CELLULAR TUMOR ANTIGEN	3.41e-03
31	61	95.3	369	13	Q9W678 CELLULAR TUMOR ANTIGEN	1.93e-02
32	59	92.2	45	13	Q92042 CELLULAR TUMOR ANTIGEN	6.00e-02
33	59	92.2	342	13	Q92143 CELLULAR TUMOR ANTIGEN	6.00e-02
34	59	92.2	342	13	O57538 CELLULAR TUMOR ANTIGEN	6.00e-02
35	59	92.2	367	13	Q9W679 CELLULAR TUMOR ANTIGEN	6.00e-02
36	59	92.2	497	11	Q9WUJ0 P73 (FRAGMENT)	6.00e-02
37	59	92.2	439	4	O15351 P73 PROTEIN	6.00e-02
38	59	92.2	636	4	O15350 P53-LIKE TRANSCRIPTION	6.00e-02
39	59	92.2	637	6	Q9XSK8 P73	6.00e-02
40	59	92.2	641	13	Q9W664 DN P63 GAMMA	1.82e-01
41	57	89.1	393	4	O75922 TA*P63 GAMMA	1.82e-01
42	57	89.1	483	11	O88897 DN P63 ALPHA	1.82e-01
43	57	89.1	586	11	O89097 KET PROTEIN (FRAGMENT)	1.82e-01
44	57	89.1	634	11	O35834 TA*P63 ALPHA	1.82e-01
45	57	89.1	680	11	O88898	1.82e-01

ALIGNMENTS

RESULT 1	PRELIMINARY;	PRT;	37 AA.
ID Q64447			
AC Q64447			
DT 01-NOV-1996 (TREMELrel. 01, Created)			
DT 01-NOV-1996 (TREMELrel. 01, Last sequence update)			
DT 01-NOV-1999 (TREMELrel. 12, Last annotation update)			
DE CELLULAR TUMOR ANTIGEN P53 (FRAGMENT)			
OS Marmota monax (Woodchuck)			
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;			
OC Eutheria; Rodentia; Sciurognathi; Sciuridae; Sciurinae; Marmota.			
RN [1]			
RP SEQUENCE FROM N.A.			
RC TISSUE=LIVER;			
RX MEDLINE; 95007566.			
RA RIVKINA M.B., CULLEN J.M., ROBINSON W.S., MARION P.L.;			
RT "State of the p53 gene in hepatocellular carcinomas of ground			
RT squirrels and woodchucks with past and ongoing infection with			
RT hepadnaviruses."			
RL Cancer Res. 54:5430-5437(1994).			
RN [2]			
RP SEQUENCE FROM N.A.			
RC TISSUE=LIVER;			
RA RIVKINA M.B., TENNANT B.C., ROBINSON W.S., MARION P.L.;			
RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.			
CC -1- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT			
CC PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL			
CC CYCLE REGULATION, IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY			
CC REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED			
CC FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF			
CC CYCLIN-DEPENDENT KINASES (BY SIMILARITY).			
CC -1- SUBCELLULAR LOCATION: NUCLEAR.			
DR EMBL; U44835; AAA86636.1; -.			
DR HSSP; P04637; ITR.			
DR PROSITE; PS00348; P53; 1.			
DR PFAM; PF00870; P53; 1.			
KW Anti-oncogene; DNA-binding; Transcription regulation; Activator;			
FT NUCLEAR protein; Phosphorylation.			
FT NON_TER 1			
FT NON_TER 37			
SQ SEQUENCE 37 AA; 4140 MW; 1EBD29B4 CRC32;			

Query Match 100.0%; Score 64; DB 11; Length 37;  
Best Local Similarity 100.0%; Pred. No. 3.41e-03; Indels 0; Gaps 0;  
Matches 9; Conservative 0; Mismatches 0;



CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
DR EMBL; U48956; AAB91534.1; -  
DR HSP; P04637; LSAH.  
DR PROSITE; PS00348; P53; 1.  
DR PFAM; PF00870; P53; 1.  
KW Anti-oncogene; DNA-binding; Transcription regulation; Activator;  
KW Nuclear protein; Phosphorylation; Apoptosis.  
FT DOMAIN 1 80 ASP/GLU-RICH (ACIDIC).  
FT DOMAIN 81 150 HYDROPHOBIC.  
FT DOMAIN 319 393 HIGHLY BASIC AND MAY BE INVOLVED IN  
INTERACTION WITH DNA.  
FT DOMAIN 311 323 NUCLEAR LOCALIZATION SIGNAL.  
FT MOD\_RES 392 392 PHOSPHORYLATION (BY SIMILARITY).  
SQ SEQUENCE 393 AA; 43655 MW; 11A9B7F8 CRC32;

Query Match 100.08; Score 64; DB 1; Length 393;  
Best Local Similarity 100.08; Pred. No. 9.32e-04;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 249 RPILTIITL 257  
QY 1 RPILTIITL 9

Search completed: Sat Apr 15 00:44:49 2000  
Job time : 42 secs.

RP VARIANT LFS LEU-272.  
RX MEDLINE; 92147883.  
RA FELIX C.A., NAU M.M., TAKAHASHI T., MITSUDOMI T., CHIBA I.,  
RA POPACK D.G., REAMAN G.H., COLE D.E., LETTERIO J.J., WHANG-PENG J.,  
RA KNUTSEN T., MINNA J.D.;  
RT "Hereditary and acquired p53 gene mutations in childhood acute  
RT lymphoblastic leukemia.";  
RL J. Clin. Invest. 89:640-647(1992).  
RN [24]  
RP VARIANTS LFS HIS-273 AND VAL-325.  
RX MEDLINE; 9228023.  
RA MALKIN D., JOLLY K.W., BARBIER N., LOOK A.T., FRIEND S.H.,  
RA GEBHARDT M.C., ANDERSEN T.I., BORRESEN A.-L., LI F.P., GARBER J.,  
RA STRONG L.C.;  
RT "Germline mutations of the p53 tumor-suppressor gene in children and  
RT young adults with second malignant neoplasms.";  
RL New Engl. J. Med. 326:1309-1315(1992).  
RN [25]  
RP VARIANTS BREAST TUMORS GLN-132; SER-249; LYS-280 AND LYS-285.  
RX MEDLINE; 90295284.  
RA BARTEK J., IGGO R., GANNON J., LANE D.P.;  
RT "Genetic and immunochemical analysis of mutant p53 in human breast  
RT cancer cell lines.";  
RL Oncogene 5:893-899(1990).  
RN [26]  
RP VARIANTS COLON TUMORS PHE-241 AND HIS-273.  
RX MEDLINE; 91017544.  
RA RODRIGUES N.R., ROWAN A., SMITH M.E.F., KERR I.B., BODMER W.F.,  
RA GANNON J.V., LANE D.P.;

... Note: remainder of annotations omitted.

Query Match 100.0%; Score 64; DB 1; Length 393;  
Best Local Similarity 100.0%; Pred. No. 9.32e-04;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 249 RPILTIITL 257  
|||||  
QY 1 RPILTIITL 9

RESULT 14  
ID P33\_CERAE STANDARD; PRT; 393 AA.  
AC P13481;  
DT 01-JAN-1990 (Rel. 13, Created)  
DT 01-JAN-1990 (Rel. 13, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE CELLULAR TUMOR ANTIGEN P53.  
GN TP53.  
OS Cercopithecus aethiops (Green monkey) (Grivet).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Primates; Catarrhini; Cercopithecoidea; Cercopithecinae;  
OC Chlorocebus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=LIVER;  
RX MEDLINE; 90045967.  
RA RIGAUDY P., ECKHART W.;  
RT "Nucleotide sequence of a cDNA encoding the monkey cellular  
RT phosphoprotein p53.";  
RL Nucleic Acids Res. 17:8375-8375(1989).  
CC -1- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES  
CC GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL  
CC CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN  
CC TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION. IT IS A  
CC TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION  
CC BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF  
CC THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.  
CC APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF  
CC BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2  
CC EXPRESSION.  
CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
CC -1- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY  
CC OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED  
CC IN MANY TYPES OF CANCER.

CC OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED  
CC IN MANY TYPES OF CANCER.  
CC -1- SIMILARITY: BELONGS TO THE P53 FAMILY.  
CC -----  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; X16384; CAA34420.1; -  
DR PIR; S06594; S06594.  
DR HSP; P04637; ISAH.  
DR PROSITE; PS00348; P53; 1.  
DR PFAM; PF00870; P53; 1.  
KW Anti-oncogene; DNA-binding; Transcription regulation; Activator;  
KW Nuclear protein; Phosphorylation; Apoptosis.  
FT DOMAIN 1 68 ASP/GLU-RICH (ACIDIC).  
FT DOMAIN 81 150 HYDROPHOBIC.  
FT DOMAIN 319 393 HIGHLY BASIC AND MAY BE INVOLVED IN  
FT INTERACTION WITH DNA.  
FT DOMAIN 311 323 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
FT MOD\_RES 392 392 PHOSPHORYLATION (BY SIMILARITY).  
SQ SEQUENCE 393 AA; 43696 MW; BBE7DC62 CRC32;  
  
Query Match 100.0%; Score 64; DB 1; Length 393;  
Best Local Similarity 100.0%; Pred. No. 9.32e-04;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Db 249 RPILTIITL 257  
|||||  
QY 1 RPILTIITL 9  
  
RESULT 15  
ID P53\_MACMU STANDARD; PRT; 393 AA.  
AC P56424;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE CELLULAR TUMOR ANTIGEN P53.  
GN TP53 OR P53.  
OS Macaca mulatta (Rhesus macaque).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Primates; Catarrhini; Cercopithecoidea; Cercopithecinae;  
OC Macaca.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA KHAN M.A., HANSEN C., WELSH J.A., BENNETT W.P.;  
RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES  
CC GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL  
CC CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN  
CC TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION. IT IS A  
CC TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION  
CC BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF  
CC THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.  
CC APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF  
CC BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2  
CC EXPRESSION.  
CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
CC -1- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY  
CC OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED  
CC IN MANY TYPES OF CANCER.

QY 1 APAPAPSWPL 10  
|||||||

## RESULT 14

ID W13972 standard; Protein: 363 AA.  
AC W13972;  
DT 25-JUN-1997 (first entry)  
DE Modified p53 variant p53Q248del364-393.  
KW p53; tumour suppressor; cancer; therapy; cell proliferation;  
KW apoptosis; protein engineering; DNA binding.  
OS Synthetic.  
PN W09710843-A1.  
PD 27-MAR-1997.  
PF 20-SEP-1996; U15188.  
PR 22-SEP-1995; US-004802.  
PR 21-AUG-1996; US-697221.  
PA (WIST-) WISTAR INST ANATOMY & BIOLOGY.  
PI Halazonetis TD;  
DR WPI; 97-202618/18.  
PT R284K modified p53 protein having DNA binding ability - useful in treatment of cancer  
PS Example 1; 53-54; 82pp; English.  
CC Modified p53 variant p53Q248del364-393 (W13972) has the tumour-derived glutamine 248 mutation (see also W13951) and a deletion of the C-terminal 30 amino acids of wild-type p53 (see also W13948). Gln248 is a Class I p53 tumour mutation that affects DNA binding. The C-terminal deletion, introduced by site-directed mutagenesis of p53 DNA, activates the DNA binding of the p53 tumour mutant. This provides the means for pharmacological rescue of p53 function in cancer patients. Other modified p53 constructs (W13949-50, W13953-54, W13968-77) have also been produced. Nucleic acids coding for modified p53 can be used for cancer gene therapy.  
SQ Sequence 363 AA;

Query Match 100.0%; Score 74; DB 1; Length 363;

Best Local Similarity 100.0%; Pred. No. 6.03e+00; Mismatches 0; Indels 0; Gaps 0;

Db 84 APAPAPSWPL 93  
|||||||

QY 1 APAPAPSWPL 10

## RESULT 15

ID W13975 standard; Protein: 363 AA.  
AC W13975;  
DT 25-JUN-1997 (first entry)  
DE Modified p53 variant p53H273R284del364-393.  
KW p53; tumour suppressor; cancer; therapy; cell proliferation;  
KW apoptosis; protein engineering; DNA binding.  
OS Synthetic.  
PN W09710843-A1.  
PD 27-MAR-1997.  
PF 20-SEP-1996; U15188.  
PR 22-SEP-1995; US-004802.  
PR 21-AUG-1996; US-697221.  
PA (WIST-) WISTAR INST ANATOMY & BIOLOGY.  
PI Halazonetis TD;  
DR WPI; 97-202618/18.  
PT R284K modified p53 protein having DNA binding ability - useful in treatment of cancer  
PS Example 1; 58-59; 82pp; English.  
CC Modified p53 variant p53H273R284del364-393 (W13975) has the tumour-derived His273 mutation (see also W13952), a Thr284 to Arg substn. (see also W13949) and a deletion of the 30 C-terminal amino acids of wild-type p53 (W13948). His273 is a Class I p53 tumour mutation that affects DNA binding. The T284R substitution, introduced by site-directed mutagenesis of p53 DNA, provides a novel p53-DNA contact between a phosphate of the DNA backbone and p53, and restores DNA binding. The C-terminal deletion permits in vitro DNA binding. The construct provides the means for pharmacological rescue of p53 function in cancer patients. Other modified p53

CC constructs (W13949-50, W13953-54, W13968-77) have also been produced. Nucleic acids coding for modified p53 can be used for cancer gene therapy.  
SQ Sequence 363 AA;  
Query Match 100.0%; Score 74; DB 1; Length 363;  
Best Local Similarity 100.0%; Pred. No. 6.03e+00; Mismatches 0; Indels 0; Gaps 0;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Db 84 APAPAPSWPL 93  
|||||||

QY 1 APAPAPSWPL 10

Search completed: Sat Apr 15 00:20:56 2000  
Job time : 54 secs.

CC acids coding for modified p53 constructs can be used for cancer  
CC gene therapy.  
SQ Sequence 359 AA;

Query Match 100.0%; Score 74; DB 1; Length 359;  
Best Local Similarity 100.0%; Pred. No. 6.03e+00;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 84 APAPAPSWPL 93  
|||||  
QY 1 APAPAPSWPL 10

RESULT 11  
ID W13961 standard; Protein; 361 AA.  
AC W13961;  
DT 25-JUN-1997 (first entry)  
DE Chimeric p53 protein.  
KW p53; tumour suppressor; cancer; therapy; cell proliferation;  
KW apoptosis; protein engineering; GCN4; DNA binding.  
OS Chimeric Homo sapiens;  
OS Chimeric synthetic.  
FH Key Location/Qualifiers  
FT region 1..323  
FT /label= p53wt  
FT /note= "amino acids 1-323 of wild-type p53"  
FT region 324..329  
FT /label= Linker  
FT region 330..361  
FT /label= GCN4  
FT /note= "amino acids 250-281 of GCN4 LZ variant"

WO9710843-A1.  
27-MAR-1997.  
PD 20-SEP-1996; U15188.  
PR 22-SEP-1995; US-004802.  
PR 21-AUG-1996; US-697221.  
PA (WIST-) WISTAR INST ANATOMY & BIOLOGY.  
PI Halazonetis TD;  
DR WPI: 97-202618/18.  
PT R284K modified p53 protein having DNA binding ability - useful in  
PT treatment of cancer  
PS Disclosure; Refer to Page 8; 82pp; English.  
CC Chimeric p53 constructs (W13956-67) comprise N-terminal portions  
CC of human wild-type p53 tumour suppressor (see also W13948) linked  
CC to a C-terminal portion of the LZ variant (see also W13955) of  
CC GCN4 and, in some cases, the C-terminal portion of wild-type  
CC p53. The chimeric proteins have DNA binding activity and can  
CC replace lost or insufficient p53 function, providing the means for  
CC pharmacological rescue of p53 function in cancer patients. Nucleic  
CC acids coding for modified p53 constructs can be used for cancer  
CC gene therapy.  
SQ Sequence 361 AA;

Query Match 100.0%; Score 74; DB 1; Length 361;  
Best Local Similarity 100.0%; Pred. No. 6.03e+00;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 84 APAPAPSWPL 93  
|||||  
QY 1 APAPAPSWPL 10

RESULT 12  
ID W28479 standard; Protein; 363 AA.  
AC W28479;  
DT 25-NOV-1997 (first entry)  
DE Human p53 protein variant V-325 encoded by pEC114.  
DE Leucine zipper domain; LZD; oligomerisation domain; mutant; mutein;  
KW substitution; replacement; transactivation; viral protein VP16; HSV;  
KW anti-oncogene; hyperproliferation; cancer; restenosis;  
KW tumour suppression; apoptosis.  
OS Chimeric - Homo sapiens.  
OS Chimeric - Herpes simplex virus.

OS Synthetic.  
PN WO9704092-A1.  
PD 06-FEB-1997.  
PF 17-JUL-1996; F01111.  
PR 19-JUL-1995; FR-008729.  
PA (RHON ) RHONE POULENC RORER SA.  
PI Bracco L, Conseiller E;  
DR WPI: 97-132633/12.  
DR N-PSDB; T86215.  
PT New p53 variants e.g. with oligomerisation domain replaced by  
PT leucine zipper - useful for treating hyper-proliferative disorders,  
PT esp. cancer and restenosis  
PS Claim 30; Pages 76-78; 133pp; French.  
CC Claimed variants of protein p53 have at least part of the  
CC oligomerisation domain deleted and replaced by a leucine zipper  
CC domain. The mutants preferably also have at least part of the p53  
CC transactivation domain (amino acids 1-74) deleted and replaced by  
CC the transactivating domain (TD) from herpes simplex virus viral  
CC protein VP16 (amino acids 411-490). The present sequence is that of  
CC a specifically claimed p53 variant designated V-325 and comprising  
CC the VP16 TD, amino acids 75-325 of human wild-type p53 and a  
CC leucine zipper domain at the C-terminal. The p53 variants are  
CC more active and more stable tumour suppressors and apoptosis-inducing  
CC agents than wild-type p53 and are active where the wild-type protein  
CC is not, i.e. they are not inactivated by dominant negative or oncogenic  
CC mutants, nor by other cellular proteins (because the leucine zipper  
CC domain prevents formation of inactive mixed oligomers).  
SQ Sequence 363 AA;

Query Match 100.0%; Score 74; DB 1; Length 363;  
Best Local Similarity 100.0%; Pred. No. 6.03e+00;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 92 APAPAPSWPL 101  
|||||  
QY 1 APAPAPSWPL 10

RESULT 13  
ID W13954 standard; Protein; 363 AA.  
AC W13954;  
DT 25-JUN-1997 (first entry)  
DE Modified p53 variant (Del364-393).  
KW p53; tumour suppressor; cancer; therapy; cell proliferation;  
KW apoptosis; protein engineering; DNA binding.  
OS Synthetic.  
PN WO9710843-A1.  
PD 27-MAR-1997.  
PF 20-SEP-1996; U15188.  
PR 22-SEP-1995; US-004802.  
PR 21-AUG-1996; US-697221.  
PA (WIST-) WISTAR INST ANATOMY & BIOLOGY.  
PI Halazonetis TD;  
DR WPI: 97-202618/18.  
PT R284K modified p53 protein having DNA binding ability - useful in  
PT treatment of cancer  
PS Example 1; 49-51; 82pp; English.  
CC A modified p53 variant (W13954) comprises wild-type p53 (see  
CC also W13948) having a deletion of the C-terminal 30 amino acids,  
CC and is obtd. by site-directed mutagenesis of p53 DNA. Deletion of  
CC the p53 C-terminal 30 amino acids activates the DNA binding of  
CC common Class I p53 mutants (see also W13951-52). Novel modified  
CC p53 variants (W13949-50, W13953-54, W13968-77), some contg.  
CC C-terminal deletions, provide the means for pharmacological rescue  
CC of p53 function in cancer patients. Nucleic acids coding for  
CC modified p53 can be used for cancer gene therapy.  
SQ Sequence 363 AA;

Query Match 100.0%; Score 74; DB 1; Length 363;  
Best Local Similarity 100.0%; Pred. No. 6.03e+00;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 84 APAPAPSWPL 93

```
Query Match      100.0%; Score 74; DB 1; Length 337;
Best Local Similarity 100.0%; Pred. No. 6.03e+00;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db      84 APAPAPSWPL 93
QY      1 APAPAPSWPL 10

RESULT
ID W28493 standard; Protein: 353 AA.
AC W28493;
DE Human p53 protein variant 393-325 encoded by p53.
KW Leucine zipper domain; LZD; oligomerisation domain; mutant; muten;
KW substitution; replacement; transactivation; viral protein VP16; HSV;
KW anti-oncogene; hyperproliferation; cancer; restenosis;
KW tumour suppression; apoptosis.
OS Homo sapiens.
OS Synthetic.
PN WO9704092-A1.
PD 06-FEB-1997.
PF 17-JUL-1996; F01111.
PR 19-JUL-1995; FR-008729.
PA (RHON ) RHONE POULENC RORER SA.
PI Bracco L, Conseiller E;
DR WPI: 97-132633/12.
DR N-PSDB: T86222.
PT New p53 variants e.g. with oligomerisation domain replaced by
PT leucine zipper - useful for treating hyper-proliferative disorders,
PT esp. cancer and restenosis.
PS Claim 37; Pages 90-92; 133pp; French.
CC Claimed variants of protein p53 have at least part of the
CC oligomerisation domain deleted and replaced by a leucine zipper
CC domain. The mutants preferably also have at least part of the p53
CC transactivation domain (amino acids 1-74) deleted and replaced by
CC a specifically claimed p53 variant designated 393-325 and comprising
CC the 325-393 of p53. The present sequence is that of
CC the 325-393 domain, amino acids 75-325 of human wild-type p53 and a
CC leucine zipper domain at the C-terminal. The p53 variants are
CC more active and more stable tumour suppressors and apoptosis-inducing
CC agents than wild-type p53 and are active where the wild-type protein
CC is not, i.e. they are not inactivated by dominant negative or oncogenic
CC mutants, nor by other cellular proteins (because the leucine zipper
CC domain prevents formation of inactive mixed oligomers).
CC Sequence 353 AA;

Query Match      100.0%; Score 74; DB 1; Length 353;
Best Local Similarity 100.0%; Pred. No. 6.03e+00;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db      82 APAPAPSWPL 91
QY      1 APAPAPSWPL 10

RESULT
ID W13960 standard; Protein: 359 AA.
AC W13960;
DE 25-JUN-1997 (first entry)
DE Chimeric p53 protein.
KW p53; tumour suppressor; cancer; therapy; cell proliferation;
KW apoptosis; protein engineering; GCN4; DNA binding.
OS Chimeric Homo sapiens;
OS Chimeric synthetic.
FH Key Location/Qualifiers
FT region 1..323
FT /label= p53wt
FT /note= "amino acids 1-323 of wild-type p53"
FT region 324..326
FT /label= Linker
FT region 327..359
FT /label= GCN4
FT /note= "amino acids 249-281 of GCN4 LZ variant"
PN WO9710843-A1.
PD 27-MAR-1997.
PF 20-SEP-1996; U15188.
PR 22-SEP-1995; US-004802.
PR 21-AUG-1996; US-697221.
PA (WIST-) WISTAR INST ANATOMY & BIOLOGY.
PI Halazonetis TD;
DR WPI: 97-202618/18.
PT R284K modified p53 protein having DNA binding ability - useful in
PT treatment of cancer
PT Disclosure; Refer to Page 8; 82pp; English.
PS Chimeric p53 constructs (W13956-67) comprise N-terminal portions
CC of human wild-type p53 tumour suppressor (see also W13948) linked
CC to a C-terminal portion of the LZ variant (see also W13955) of
CC GCN4 and, in some cases, the C-terminal portion of wild-type
CC p53. The chimeric proteins have DNA binding activity and can
CC replace lost or insufficient p53 function, providing the means for
CC pharmacological rescue of p53 function in cancer patients. Nucleic
```

```
PD 06-FEB-1997.
PF 17-JUL-1996; F01111.
PR 19-JUL-1995; FR-008729.
PA (RHON ) RHONE POULENC RORER SA.
PI Bracco L, Conseiller E;
DR WPI: 97-132633/12.
PT New p53 variants e.g. with oligomerisation domain replaced by
PT leucine zipper - useful for treating hyper-proliferative disorders,
PT esp. cancer and restenosis.
PS Claim 37; Pages 90-92; 133pp; French.
CC Claimed variants of protein p53 have at least part of the
CC oligomerisation domain deleted and replaced by a leucine zipper
CC domain. The mutants preferably also have at least part of the p53
CC transactivation domain (amino acids 1-74) deleted and replaced by
CC the domain 325-393 of p53. The present sequence is that of a
CC specifically claimed p53 variant designated 393-325H and comprising
CC the 325-393 domain, amino acids 75-325 of human wild-type p53 (but with
CC Arg182 replaced by His) and a leucine zipper domain at the C-terminal.
CC The p53 variants are more active and more stable tumour suppressors
CC and apoptosis-inducing agents than wild-type p53 and are active where
CC the wild-type protein is not, i.e. they are not inactivated by dominant
CC negative or oncogenic mutants, nor by other cellular proteins (because
CC the leucine zipper domain prevents formation of inactive mixed
CC oligomers).
CC (Note: this sequence does not appear in the specification and has
CC been produced by modifying the given sequence of variant 393-325).
CC Sequence 353 AA;

Query Match      100.0%; Score 74; DB 1; Length 353;
Best Local Similarity 100.0%; Pred. No. 6.03e+00;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db      82 APAPAPSWPL 91
QY      1 APAPAPSWPL 10

RESULT
ID W13960 standard; Protein: 359 AA.
AC W13960;
DE 25-JUN-1997 (first entry)
DE Chimeric p53 protein.
KW p53; tumour suppressor; cancer; therapy; cell proliferation;
KW apoptosis; protein engineering; GCN4; DNA binding.
OS Chimeric Homo sapiens;
OS Chimeric synthetic.
FH Key Location/Qualifiers
FT region 1..323
FT /label= p53wt
FT /note= "amino acids 1-323 of wild-type p53"
FT region 324..326
FT /label= Linker
FT region 327..359
FT /label= GCN4
FT /note= "amino acids 249-281 of GCN4 LZ variant"
PN WO9710843-A1.
PD 27-MAR-1997.
PF 20-SEP-1996; U15188.
PR 22-SEP-1995; US-004802.
PR 21-AUG-1996; US-697221.
PA (WIST-) WISTAR INST ANATOMY & BIOLOGY.
PI Halazonetis TD;
DR WPI: 97-202618/18.
PT R284K modified p53 protein having DNA binding ability - useful in
PT treatment of cancer
PT Disclosure; Refer to Page 8; 82pp; English.
PS Chimeric p53 constructs (W13956-67) comprise N-terminal portions
CC of human wild-type p53 tumour suppressor (see also W13948) linked
CC to a C-terminal portion of the LZ variant (see also W13955) of
CC GCN4 and, in some cases, the C-terminal portion of wild-type
CC p53. The chimeric proteins have DNA binding activity and can
CC replace lost or insufficient p53 function, providing the means for
CC pharmacological rescue of p53 function in cancer patients. Nucleic
```

```

RESULT 5
ID W28498 standard; Protein; 335 AA.
AC W28498;
DE Human p53 protein variant 360h-325H.
KW Leucine zipper domain; LZD; oligomerisation domain; mutant; muten;
KW substitution; replacement; transactivation; hinge region;
KW anti-oncogene; hyperproliferation; cancer; restenosis;
KW tumour suppression; apoptosis.
OS Homo sapiens.
OS Synthetic.
FH Key Location/Qualifiers
FT region 39..53 /label= hinge
FT misc_difference 161
FT /note= "Arg residue at position 182 of wild-type
FT p53 has been mutated to His"
FN W09704092-A1.
PD 06-FEB-1997.
PF 17-JUL-1996; F01111.
PR 19-JUL-1995; FR-008729.
PA (RHON ) RHONE POULENC RORER SA.
PI Bracco L, Conseiller E;
DR WPI; 97-132633/12.
PT New p53 variants e.g. with oligomerisation domain replaced by
PT leucine zipper - useful for treating hyper-proliferative disorders,
PT esp. cancer and restenosis
PS Claim 39; Page -; 133pp; French.
CC Claimed variants of protein p53 have at least part of the
CC oligomerisation domain deleted and replaced by a leucine zipper
CC domain. The mutants preferably also have at least part of the p53
CC transactivation domain (amino acids 1-74) deleted and replaced by
CC the domain 325-360 of p53. The present sequence is that of a
CC specifically claimed p53 variant designated 360h-325H and comprising
CC the 325-360 domain, separated from amino acids 75-325 of human
CC wild-type p53 by a synthetic hinge sequence (Gly4Ser)3 and with a
CC leucine zipper domain at the C-terminal. The p53 variants are
CC more active and more stable tumour suppressors and apoptosis-inducing
CC agents than wild-type p53 and are active where the wild-type protein
CC is not, i.e. they are not inactivated by dominant negative or oncogenic
CC mutants, nor by other cellular proteins (because the leucine zipper
CC domain prevents formation of inactive mixed oligomers).
SQ Sequence 335 AA;

Query Match 100.0%; Score 74; DB 1; Length 335;
Best Local Similarity 100.0%; Pred. No. 6.03e+00;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 64 APAPAPSWPL 73
| | | | | | | | | |
QY 1 APAPAPSWPL 10

RESULT 7
ID W13962 standard; Protein; 337 AA.
AC W13962;
DE 25-JUN-1997 (first entry)
DE Chimeric p53 protein.
KW p53; tumour suppressor; cancer; therapy; cell proliferation;
KW apoptosis; protein engineering; GCN4; DNA binding.
OS Chimeric Homo sapiens;
OS Chimeric synthetic.
FH Key Location/Qualifiers
FT region 1..300 /label= p53wt
FT /note= "amino acids 1-300 of wild-type p53"
FT region 301..305 /label= Linker
FT region 306..337 /label= GCN4
FT /note= "amino acids 250-281 of GCN4 LZ variant"
FN W09710843-A1.
PD 27-MAR-1997.
PF 20-SEP-1996; U15188.
PR 22-SEP-1995; US-004802.
PR 21-AUG-1996; US-697221.
PA (WIST-) WISTAR INST ANATOMY & BIOLOGY.
PI Halazonetis TD;
DR WPI; 97-202618/18.
PT R284K modified p53 protein having DNA binding ability - useful in
PT treatment of cancer
PS Disclosure: Refer to Page 8; 82pp; English.
CC Chimeric p53 constructs (W13956-57) comprise N-terminal portions
CC of human wild-type p53 tumour suppressor (see also W13948) linked
CC to a C-terminal portion of the LZ variant (see also W13955) of
CC GCN4 and, in some cases, the C-terminal portion of wild-type
CC p53. The chimeric proteins have DNA binding activity and can
CC replace lost or insufficient p53 function, providing the means for
CC pharmacological rescue of p53 function in cancer patients. Nucleic
CC acids coding for modified p53 constructs can be used for cancer
CC gene therapy.
SQ Sequence 337 AA;

```

```

PR 19-JUL-1995; FR-008729.
PA (RHON ) RHONE POULENC RORER SA.
PI Bracco L, Conseiller E;
DR WPI; 97-132633/12.
DR N-PSDB; T86224.
PT New p53 variants e.g. with oligomerisation domain replaced by
PT leucine zipper - useful for treating hyper-proliferative disorders,
PT esp. cancer and restenosis
PS Claim 39; Pages 94-95; 133pp; French.
CC Claimed variants of protein p53 have at least part of the
CC oligomerisation domain deleted and replaced by a leucine zipper
CC domain. The mutants preferably also have at least part of the p53
CC transactivation domain (amino acids 1-74) deleted and replaced by
CC the domain 325-360 of p53. The present sequence is that of a
CC specifically claimed p53 variant designated 360h-325 and comprising
CC the 325-360 domain, separated from amino acids 75-325 of human
CC wild-type p53 by a synthetic hinge sequence (Gly4Ser)3 and with a
CC leucine zipper domain at the C-terminal. The p53 variants are
CC more active and more stable tumour suppressors and apoptosis-inducing
CC agents than wild-type p53 and are active where the wild-type protein
CC is not, i.e. they are not inactivated by dominant negative or oncogenic
CC mutants, nor by other cellular proteins (because the leucine zipper
CC domain prevents formation of inactive mixed oligomers).
SQ Sequence 335 AA;

Query Match 100.0%; Score 74; DB 1; Length 335;
Best Local Similarity 100.0%; Pred. No. 6.03e+00;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 64 APAPAPSWPL 73
| | | | | | | | | |
QY 1 APAPAPSWPL 10

RESULT 7
ID W13962 standard; Protein; 337 AA.
AC W13962;
DE 25-JUN-1997 (first entry)
DE Chimeric p53 protein.
KW p53; tumour suppressor; cancer; therapy; cell proliferation;
KW apoptosis; protein engineering; GCN4; DNA binding.
OS Chimeric Homo sapiens;
OS Chimeric synthetic.
FH Key Location/Qualifiers
FT region 1..300 /label= p53wt
FT /note= "amino acids 1-300 of wild-type p53"
FT region 301..305 /label= Linker
FT region 306..337 /label= GCN4
FT /note= "amino acids 250-281 of GCN4 LZ variant"
FN W09710843-A1.
PD 27-MAR-1997.
PF 20-SEP-1996; U15188.
PR 22-SEP-1995; US-004802.
PR 21-AUG-1996; US-697221.
PA (WIST-) WISTAR INST ANATOMY & BIOLOGY.
PI Halazonetis TD;
DR WPI; 97-202618/18.
PT R284K modified p53 protein having DNA binding ability - useful in
PT treatment of cancer
PS Disclosure: Refer to Page 8; 82pp; English.
CC Chimeric p53 constructs (W13956-57) comprise N-terminal portions
CC of human wild-type p53 tumour suppressor (see also W13948) linked
CC to a C-terminal portion of the LZ variant (see also W13955) of
CC GCN4 and, in some cases, the C-terminal portion of wild-type
CC p53. The chimeric proteins have DNA binding activity and can
CC replace lost or insufficient p53 function, providing the means for
CC pharmacological rescue of p53 function in cancer patients. Nucleic
CC acids coding for modified p53 constructs can be used for cancer
CC gene therapy.
SQ Sequence 337 AA;

```

RESULT 2  
ID W28483 standard; Protein; 253 AA.  
AC W28483;  
DT 25-NOV-1997 (first entry)  
DE Human p53 protein variant v-367 encoded by pEC141.  
KW Leucine zipper domain; LZD; oligomerisation domain; mutant; muten;  
KW substitution; replacement; transactivation; viral protein VP16; HSV;  
KW anti-oncogene; hyperproliferation; cancer; restenosis;  
KW tumour suppression; apoptosis.  
OS Chimeric - Homo sapiens.  
OS Synthetic.  
PN WO9704092-A1.  
PD 06-FEB-1997.  
PF 17-JUL-1996; F01111.  
PR 19-JUL-1995; FR-008729.  
PA (RHON ) RHONE POULENC RORER SA.  
PI Bracco L, Conseiller E;  
DR N-PSDB; T86217.  
PT New p53 variants e.g. with oligomerisation domain replaced by  
PT leucine zipper - useful for treating hyper-proliferative disorders,  
PT esp. cancer and restenosis  
PS Claim 32; Pages 80-81; 133pp; French.  
CC Claimed variants of protein p53 have at least part of the p53  
CC transactivation domain (amino acids 1-74) deleted and replaced by  
CC the transactivating domain (TD) from herpes simplex virus viral  
CC protein VP16 (amino acids 411-490). The present sequence is that of  
CC a specifically claimed p53 variant designated v-367 and comprising  
CC the VP16 TD with amino acids 75-367 of human wild-type p53. The p53  
CC variants are more active and more stable tumour suppressors and  
CC apoptosis-inducing agents than wild-type p53 and are active where  
CC the wild-type protein is not.  
SQ Sequence 253 AA;

Query Match 100.0%; Score 74; DB 1; Length 253;  
Best Local Similarity 100.0%; Pred. No. 6.03e+00;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 92 APAPAPSWPL 101  
QY 1 APAPAPSWPL 10  
|||||

RESULT 3  
ID W28495 standard; Protein; 319 AA.  
AC W28495;  
DT 25-NOV-1997 (first entry)  
DE Human p53 protein variant 360-325 encoded by pEC178.  
KW Leucine zipper domain; LZD; oligomerisation domain; mutant; muten;  
KW substitution; replacement; transactivation; viral protein VP16; HSV;  
KW anti-oncogene; hyperproliferation; cancer; restenosis;  
KW tumour suppression; apoptosis.  
OS Homo sapiens.  
OS Synthetic.  
PN WO9704092-A1.  
PD 06-FEB-1997.  
PF 17-JUL-1996; F01111.  
PR 19-JUL-1995; FR-008729.  
PA (RHON ) RHONE POULENC RORER SA.  
PI Bracco L, Conseiller E;  
DR N-PSDB; T86223.  
PT New p53 variants e.g. with oligomerisation domain replaced by  
PT leucine zipper - useful for treating hyper-proliferative disorders,  
PT esp. cancer and restenosis  
PS Claim 38; Pages 92-94; 133pp; French.  
CC Claimed variants of protein p53 have at least part of the  
CC oligomerisation domain deleted and replaced by a leucine zipper  
CC domain. The mutants preferably also have at least part of the p53  
CC transactivation domain (amino acids 1-74) deleted and replaced by

CC the domain 325-360 of p53. The present sequence is that of a  
CC specifically claimed p53 variant designated 360-325 and comprising  
CC the 325-360 domain, amino acids 75-325 of human wild-type p53 and a  
CC leucine zipper domain at the C-terminal. The p53 variants are  
CC more active and more stable tumour suppressors and apoptosis-inducing  
CC agents than wild-type p53 and are active where the wild-type protein  
CC is not, i.e. they are not inactivated by dominant negative or oncogenic  
CC mutants, nor by other cellular proteins (because the leucine zipper  
CC domain prevents formation of inactive mixed oligomers).  
SQ Sequence 319 AA;

Query Match 100.0%; Score 74; DB 1; Length 319;  
Best Local Similarity 100.0%; Pred. No. 6.03e+00;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 48 APAPAPSWPL 57  
QY 1 APAPAPSWPL 10  
|||||

RESULT 4  
ID W28496 standard; Protein; 319 AA.  
AC W28496;  
DT 25-NOV-1997 (first entry)  
DE Human p53 protein variant 360-325H.  
KW Leucine zipper domain; LZD; oligomerisation domain; mutant; muten;  
KW substitution; replacement; transactivation; viral protein VP16; HSV;  
KW anti-oncogene; hyperproliferation; cancer; restenosis;  
KW tumour suppression; apoptosis.  
OS Homo sapiens.  
OS Synthetic.  
FH Key Location/Qualifiers  
FT misc\_difference 145  
FT /note= "Arg residue at position 182 of wild-type  
FT p53 has been mutated to His"  
PN WO9704092-A1.  
PD 06-FEB-1997.  
PF 17-JUL-1996; F01111.  
PR 19-JUL-1995; FR-008729.  
PA (RHON ) RHONE POULENC RORER SA.  
PI Bracco L, Conseiller E;  
DR WPI; 97-132633/12.  
PT New p53 variants e.g. with oligomerisation domain replaced by  
PT leucine zipper - useful for treating hyper-proliferative disorders,  
PT esp. cancer and restenosis  
PS Claim 38; Page -; 133pp; French.  
CC Claimed variants of protein p53 have at least part of the  
CC oligomerisation domain deleted and replaced by a leucine zipper  
CC domain. The mutants preferably also have at least part of the p53  
CC transactivation domain (amino acids 1-74) deleted and replaced by  
CC the domain 325-360 of p53. The present sequence is that of a  
CC specifically claimed p53 variant designated 360-325H and comprising  
CC the 325-360 domain, amino acids 75-325 of human wild-type p53 (but with  
CC Arg182 replaced by His) and a leucine zipper domain at the C-terminal.  
CC The p53 variants are more active and more stable tumour suppressors  
CC and apoptosis-inducing agents than wild-type p53 and are active where  
CC the wild-type protein is not, i.e. they are not inactivated by dominant  
CC negative or oncogenic mutants, nor by other cellular proteins (because  
CC the leucine zipper domain prevents formation of inactive mixed  
CC oligomers).  
CC (Note: this sequence does not appear in the specification and has  
CC been produced by modifying the given sequence of variant 360-325).  
SQ Sequence 319 AA;

Query Match 100.0%; Score 74; DB 1; Length 319;  
Best Local Similarity 100.0%; Pred. No. 6.03e+00;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 48 APAPAPSWPL 57  
QY 1 APAPAPSWPL 10  
|||||

\*\*\*\*\*  
W P S R L  
\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Sat Apr 15 00:20:02 2000; MasPar time 5.68 Seconds  
Tabular output not generated. 41.727 Million cell updates/sec

Title: >US-08-452-843-14  
Description: (1-10) from US08452843.pep  
Perfect Score: 74  
Sequence: 1 APAPAPSWPL 10

Scoring table: PAM 150  
Gap 15

Searched: 188963 seqs, 23686106 residues

Post-processing: Minimum Match 0.8  
Listing first 45 summaries

Database: a-geneseq36  
l-geneseqp

Statistics: Mean 16.556; Variance 72.913; scale 0.227

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	74	100.0	253	1 W28484	Human p53 protein vari	6.03e+00
2	74	100.0	253	1 W28483	Human p53 protein vari	6.03e+00
3	74	100.0	319	1 W28495	Human p53 protein vari	6.03e+00
4	74	100.0	319	1 W28496	Human p53 protein vari	6.03e+00
5	74	100.0	335	1 W28498	Human p53 protein vari	6.03e+00
6	74	100.0	335	1 W28497	Human p53 protein vari	6.03e+00
7	74	100.0	337	1 W13962	Chimeric p53 protein.	6.03e+00
8	74	100.0	353	1 W28493	Human p53 protein vari	6.03e+00
9	74	100.0	353	1 W28494	Human p53 protein vari	6.03e+00
10	74	100.0	359	1 W13960	Chimeric p53 protein.	6.03e+00
11	74	100.0	361	1 W13961	Chimeric p53 protein.	6.03e+00
12	74	100.0	363	1 W28479	Human p53 protein vari	6.03e+00
13	74	100.0	363	1 W13954	Modified p53 variant (	6.03e+00
14	74	100.0	363	1 W13972	Modified p53 variant p	6.03e+00
15	74	100.0	363	1 W13975	Modified p53 variant p	6.03e+00
16	74	100.0	363	1 W13976	Modified p53 variant p	6.03e+00
17	74	100.0	363	1 W13971	Modified p53 variant p	6.03e+00
18	74	100.0	363	1 W28480	Human p53 protein vari	6.03e+00
19	74	100.0	374	1 W28482	Human p53 protein vari	6.03e+00
20	74	100.0	374	1 W28481	Human p53 protein vari	6.03e+00
21	74	100.0	381	1 W28489	Human p53 protein vari	6.03e+00
22	74	100.0	393	1 Y03191	Amino acid sequence of	6.03e+00
23	74	100.0	393	1 W84270	Human p53 protein.	6.03e+00

24	74	100.0	393	1 W69218	Human p53 mutant 1.	6.03e+00
25	74	100.0	393	1 W69217	Human wild-type p53 pr	6.03e+00
26	74	100.0	393	1 W57245	Human p53 protein SEQ	6.03e+00
27	74	100.0	393	1 W57244	Human p53 protein SEQ	6.03e+00
28	74	100.0	393	1 W05346	Human p53 mutant R273H	6.03e+00
29	74	100.0	393	1 W05347	Human p53 mutant R248Q	6.03e+00
30	74	100.0	393	1 W13968	Modified p53 variant p	6.03e+00
31	74	100.0	393	1 W13970	Modified p53 variant p	6.03e+00
32	74	100.0	393	1 W25155	Human p53 variant foun	6.03e+00
33	74	100.0	393	1 W05349	Human p53 mutant R273C	6.03e+00
34	74	100.0	393	1 R91933	Wild type p53 protein.	6.03e+00
35	74	100.0	393	1 W05348	Human p53 mutant R282W	6.03e+00
36	74	100.0	393	1 W02617	Human p53 tumour suppr	6.03e+00
37	74	100.0	393	1 W13978	Human tumour-derived p	6.03e+00
38	74	100.0	393	1 W13952	Human tumour-derived p	6.03e+00
39	74	100.0	393	1 W13951	Human tumour-derived p	6.03e+00
40	74	100.0	393	1 W13949	T284R modified human p	6.03e+00
41	74	100.0	401	1 W28488	Human p53 protein vari	6.03e+00
42	74	100.0	402	1 W13965	Chimeric p53 protein.	6.03e+00
43	74	100.0	406	1 W13966	Chimeric p53 protein.	6.03e+00
44	74	100.0	411	1 W13967	Chimeric p53 protein.	6.03e+00
45	74	100.0	535	1 W28491	Human p53 protein vari	6.03e+00

ALIGNMENTS

RESULT 1  
ID W28484 standard; Protein; 253 AA.  
AC W28484; 253 AA.  
DT 25-NOV-1997 (first entry)  
DE Human p53 protein variant V-367H.  
KW Leucine zipper domain; L2D; Oligomerisation domain; mutant; mutein;  
KW substitution; replacement; transactivation; viral protein VP16; HSV;  
KW anti-oncogene; hyperproliferation; cancer; restenosis;  
KW tumour suppression; apoptosis.  
OS Chimeric - Homo sapiens.  
OS Chimeric - Herpes simplex virus.  
OS Synthetic.  
FH Key Location/Qualifiers  
FT misc\_difference 189  
FT /note= "Arg residue at position 182 of wild-type p53 has been mutated to His"

WO9704092-A1.  
06-FEB-1997.  
17-JUL-1996; F01111.  
19-JUL-1995; FR-008729.  
PR (RHON) RHONE FOULENC RORER SA.  
PA Bracco L, Conseiller E;  
PI WPI; 97-132633/12.  
DR New p53 variants e.g. with oligomerisation domain replaced by leucine zipper - useful for treating hyper-proliferative disorders, esp. cancer and restenosis  
PT Claim 32; Page -; 133pp; French.  
PS Claimed variants of protein p53 have at least part of the p53 transactivation domain (amino acids 1-74) deleted and replaced by the transactivating domain (TD) from herpes simplex virus viral protein VP16 (amino acids 411-490). The present sequence is that of a specifically claimed p53 variant designated V-367 and comprising the VP16 TD and amino acids 75-367 of human wild-type p53 (but with Arg182 replaced by His). The p53 variants are more active and more stable tumour suppressors and apoptosis-inducing agents than wild-type p53 and are active where the wild-type protein is not.  
CC (Note: this sequence does not appear in the specification and has been produced by modifying the given sequence of variant V-367).  
SQ Sequence 253 AA;

Query Match 100.0%; Score 74; DB 1; Length 253;  
Best Local Similarity 100.0%; Pred. No. 6.03e+00;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

D0 92 APAPAPSWPL 101  
QY 1 APAPAPSWPL 10



THE  
STAMP (used)

Best Local Similarity 55.6%; Pred. No. 8.29e+00; Mismatches 3; Indels 0; Gaps 0;

Db 2187 FSLPKFYLL 2195  
QY 1 FAMPNEYTL 9

RESULT 13  
ID O01397 PRELIMINARY; PRT; 2802 AA.  
AC O01397;  
DT 01-JUL-1997 (TREMELrel. 04, Created)  
DT 01-JUL-1997 (TREMELrel. 04, Last sequence update)  
DT 01-NOV-1998 (TREMELrel. 08, Last annotation update)  
DE NEUROFIBROMIN.  
GN NFI.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-CANTON S.  
RA HANNIGAN G.E., THE I., SHAMANSKI F.L., ORR-WEAVER T.L., GUSELLA J.F.,  
RA BERNARDS A.;  
RL Science 0:0-0(0).  
DR EMBL; L26500; AAB58977.1; -.  
DR FLYBASE; FBgn0015269; Nfl.  
DR PFAM; PF00616; RasGAP; 1.  
SQ SEQUENCE 2802 AA; 317203 MW; 032CE079 CRC32;

Query Match 72.0%; Score 54; DB 5; Length 2802;  
Best Local Similarity 55.6%; Pred. No. 8.29e+00;  
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 2187 FSLPKFYLL 2195  
QY 1 FAMPNEYTL 9

RESULT 14  
ID O01398 PRELIMINARY; PRT; 2802 AA.  
AC O01398;  
DT 01-JUL-1997 (TREMELrel. 04, Created)  
DT 01-JUL-1997 (TREMELrel. 04, Last sequence update)  
DT 01-NOV-1998 (TREMELrel. 08, Last annotation update)  
DE NEUROFIBROMIN.  
GN NFI.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-CANTON S.  
RA THE I., HANNIGAN G.E., COWLEY G.S., REGINALD S., ZHONG Y.,  
RA GUSELLA J.F., HARIHARAN I.K., BERNARDS A.;  
RT Rescue of a Drosophila Nfl mutant phenotype by protein kinase A.;  
RL Science 276:791-794(1997).  
DR EMBL; L26501; AAB58975.1; -.  
DR FLYBASE; FBgn0015269; Nfl.  
DR PFAM; PF00616; RasGAP; 1.  
SQ SEQUENCE 2802 AA; 317210 MW; 76822162 CRC32;

Query Match 72.0%; Score 54; DB 5; Length 2802;  
Best Local Similarity 55.6%; Pred. No. 8.29e+00;  
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 2187 FSLPKFYLL 2195  
QY 1 FAMPNEYTL 9

RESULT 15

ID P97526 PRELIMINARY; PRT; 2820 AA.  
AC P97526;  
DT 01-MAY-1997 (TREMELrel. 03, Created)  
DT 01-MAY-1997 (TREMELrel. 03, Last sequence update)  
DT 01-NOV-1999 (TREMELrel. 12, Last annotation update)  
DE NEUROFIBROMIN.  
GN NFI.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-WISTAR; TISSUE-BRAIN;  
RA KYRITSIS A.P., LEE P.S., MOCHIZUKI H., NISHI T., LEVIN V.A., SAYA H.;  
RT "Differential splicing of the neurofibromatosis type 1 (Nf1) gene in  
rats: Homologous splice variants in human are expressed in rat  
cells.";  
RT cells.";  
RL Int. J. Oncol. 1:149-152(1992).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-WISTAR; TISSUE-BRAIN;  
RX MEDLINE; 97137538.  
RA SUZUKI H., TAKAHASHI K., YASUMOTO K., FUSE N., SHIBAHARA S.;  
RT "Differential tissue-specific expression of neurofibromin isoform  
mRNAs in rat.";  
RL J. Biochem. 120:1048-1054(1996).  
DR EMBL; D45201; BAA08141.1; -.  
DR PROSITE; PS00509; RAS\_GTPASE\_ACTIV\_1; 1.  
DR PFAM; PF00616; RasGAP; 1.  
SQ SEQUENCE 2820 AA; 317079 MW; 6470B267 CRC32;

Query Match 72.0%; Score 54; DB 11; Length 2820;  
Best Local Similarity 55.6%; Pred. No. 8.29e+00;  
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 2137 FSLPKFYLL 2145  
QY 1 FAMPNEYTL 9

Search completed: Fri Apr 14 23:34:24 2000  
Job time : 105 secs.

Query Match 72.0%; Score 54; DB 1; Length 206;  
 Best Local Similarity 44.4%; Pred. No. 8.29e+00;  
 Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 14 FSNATYSL 22  
 QY 1 FAMPNFYTL 9  
 |:::|::|

RESULT 9  
 ID O05943 PRELIMINARY; PRT; 339 AA.  
 AC O05943;  
 DT 01-JUL-1997 (TREMBLrel. 04, Created)  
 DT 01-AUG-1999 (TREMBLrel. 11, Last sequence update)  
 DT 01-AUG-1999 (TREMBLrel. 11, Last annotation update)  
 DE CYTOCHROME D UBIQUINOL OXIDASE SUBUNIT II (CYDB) (CYTOCHROME OXIDASE  
 D, SUBUNIT II).  
 GN RP217 OR CYDB.  
 OS Rickettsia prowazekii.  
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
 OC Rickettsiaceae; Rickettsiae; Rickettsia.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-MADRID E;  
 RX MEDLINE; 99039499.  
 RA ANDERSSON S.G.E., ZOMORODIPOUR A., ANDERSSON J.O.,  
 RA SICHERITZ-PONTEN T., ALSMARK U.C.M., PODOWSKI R.M., NAEGLUND A.K.,  
 RA ERIKSSON A.S., WINKLER H.H., KURLAND C.G.;  
 RT "The genome sequence of Rickettsia prowazekii and the origin of  
 RT mitochondria";  
 RL Nature 396:133-140(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-MADRID E;  
 RX MEDLINE; 97419517.  
 RA ANDERSSON J.O., ANDERSSON S.G.E.;  
 RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE OF 53-339 FROM N.A.  
 RC STRAIN-MADRID E;  
 RX MEDLINE; 97419517.  
 RA "Genomic rearrangements during evolution of the obligate intracellular  
 RT parasite Rickettsia prowazekii as inferred from an analysis of 52015  
 RT bp nucleotide sequence";  
 RL Microbiology 143:2783-2795(1997).  
 DR EMBL; AJ235270; CAA14680.1; -.  
 DR EMBL; Y11780; CAA72465.1; -.  
 SQ SEQUENCE 339 AA; 38029 MW; F1AE57CD CRC32;

Query Match 72.0%; Score 54; DB 2; Length 339;  
 Best Local Similarity 66.7%; Pred. No. 8.29e+00;  
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 230 FSNPNYYL 238  
 QY 1 FAMPNFYTL 9  
 |:::|::|

RESULT 10  
 ID O17451 PRELIMINARY; PRT; 466 AA.  
 AC O17451;  
 DT 01-JAN-1998 (TREMBLrel. 05, Created)  
 DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)  
 DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)  
 DE GAG-LIKE PROTEIN.  
 GN GAG.  
 OS Culex pipiens (House mosquito).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;  
 OC Culicidae; Culex.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA BENSADI-MERCHERMEK N., CAGNON C., DESMONS I., SALVADO J.C.,

RA KARAMA S., D'AMICO F., MOUCHES C.;  
 RL Genetica 0:0-0(1997).  
 DR EMBL; AF030588; AAB86424.1; -.  
 DR PRINTS; PR00939; C2HCZNFINGER.  
 SQ SEQUENCE 466 AA; 51269 MW; 270BA37 CRC32;

Query Match 72.0%; Score 54; DB 5; Length 466;  
 Best Local Similarity 44.4%; Pred. No. 8.29e+00;  
 Matches 4; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

Db 426 FTLPEFFAL 434  
 QY 1 FAMPNFYTL 9  
 |:::|::|

RESULT 11  
 ID O9YGV2 PRELIMINARY; PRT; 2763 AA.  
 AC O9YGV2;  
 DT 01-MAY-1999 (TREMBLrel. 10, Created)  
 DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)  
 DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)  
 DE NEUROFIBROMATOSIS TYPE 1.  
 GN NF1.  
 OS Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Actinopterygii;  
 OC Neopterygii; Teleostei; Euteleostei; Acanthopterygii; Percomorpha;  
 OC Tetraodontiformes; Tetraodontidae; Tetraodontidae; Fugu.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 99033011.  
 RA KEHRER-SAWATZKI H., MAIER C., MOSCHGATH E., ELGAR G., KRONE W.;  
 RT "Genomic characterization of the Neurofibromatosis Type 1 gene of Fugu  
 RL Gene 222:145-153(1998).  
 DR EMBL; AF064564; AAD15839.1; -.  
 DR PROSITE; PS00509; RAS\_GTPASE\_ACTIV.1; 1.  
 SQ SEQUENCE 2763 AA; 31101 MW; 73905228 CRC32;

Query Match 72.0%; Score 54; DB 13; Length 2763;  
 Best Local Similarity 55.6%; Pred. No. 8.29e+00;  
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 2085 FSLPFXLL 2093  
 QY 1 FAMPNFYTL 9  
 |:::|::|

RESULT 12  
 ID O01399 PRELIMINARY; PRT; 2764 AA.  
 AC O01399;  
 DT 01-JUL-1997 (TREMBLrel. 04, Created)  
 DT 01-JUL-1997 (TREMBLrel. 04, Last sequence update)  
 DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)  
 DE NEUROFIBROMIN.  
 GN NF1.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CANTON S;  
 RX MEDLINE; 97277221.  
 RA THE I., HANNIGAN G.E., COWLEY G.S., REGINALD S., ZHONG Y.,  
 RA GUSELLA J.F., HARIHARAN I.K., BERNARDS A.;  
 RT "Rescue of a Drosophila NF1 mutant phenotype by protein kinase A";  
 RL Science 276:791-794(1997).  
 DR EMBL; L26502; AAB58976.1; -.  
 DR FLYBASE; FBgn0015269; NF1.  
 DR PFAM; PF00616; RasGAP; 1.  
 SQ SEQUENCE 2764 AA; 312936 MW; 54B6B40F CRC32;

Query Match 72.0%; Score 54; DB 5; Length 2764;

RA GUISEPPI G., GUY B.J., HAGA K., HAITECH J., HARWOOD C.R., HENAUT A.,  
 RA HILBERT H., HOLSAPPEL S., HOSONO S., HULLO M.F., ITAYA M., JONES L.,  
 RA JORIS B., KARAMATA D., KASAHARA Y., KLAERR-BLANCHARD M., KLEIN C.,  
 RA KOBAYASHI Y., KOETTER P., KONINGSTEIN G., KROGH S., KUMANO M.,  
 RA KURIYA K., LAPIDUS A., LARDINOIS S., LAUBER J., LAZAREVIC V.,  
 RA LEE S.M., LEVINE A., LIU H., MASUDA S., MAUEL C., MEDIGUE C.,  
 RA MEDINA N., MELLADO R.P., MIZUNO M., MOESTL D., NAKAI S., NOBACK M.,  
 RA NOONE D., O'REILLY M., OGAWA K., OGIWARA A., OUDEGA B., PARK S.H.,  
 RA PARRO V., POHL T.M., PORTELELLA D., PORMOLLIK S., PRESCOTT A.M.,  
 RA PRESECAN E., PUJIC P., PUENELLE B., RAPOPORT G., REY M., REYNOLDS S.,  
 RA RIEGER M., RIVOLTA C., ROCHA E., ROCHE B., ROSE M., SADAIE Y.,  
 RA SATO T., SCANLAN E., SCHLEICH S., SCHROETER R., SCOFFONE F.,  
 RA SEKIGUCHI J., SEKOWSKA A., SERO S.J., SERROR P., SHIN B.S., SOLDI B.,  
 RA SOROKIN A., TACCONE E., TAKAGI T., TAKAHASHI H., TAKEMARU K.,  
 RA TAKEUCHI M., TAKAKOSHI A., TANAKA T., TERPSTRA P., TOGNONI A.,  
 RA TOSATO V., UCHIYAMA S., VANDENBOL M., VANNIER F., VASSAROTTI A.,  
 RA VIARI A., WAMBUIT R., WEDLER E., WEDLER H., WEITZENEGGER T.,  
 RA WINTERS P., WIPAT A., YAMAMOTO H., YAMANE K., YASUMOTO K., YATA K.,  
 RA YOSHIDA K., YOSHIKAWA H.F., ZUMSTEIN E., YOSHIKAWA H., DANCHIN A.,  
 RT "The complete genome sequence of the gram-positive bacterium *Bacillus*  
 RT *subtilis*.";  
 RL Nature 390:249-256(1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-168;  
 RA KUNST F., OGASAWARA N., YOSHIKAWA H., DANCHIN A.;  
 RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE OF 126-786 FROM N.A.  
 RC STRAIN-168;  
 RA GHIM S.-Y., CHOI S.-K., SHIN B.-S., PARK S.-H.;  
 RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; Z99114; CAB13897.1; -;  
 DR EMBL; AF012906; AAB92484.1; -;  
 DR PRAM; PF00317; ribonucleo-red; 2;  
 SQ SEQUENCE 786 AA; 90101 MW; AD87D8C8 CRC32;

Query Match 73.3%; Score 55; DB 2; Length 786;  
 Best Local Similarity 66.7%; Pred. No. 5.30e+00;  
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 578 FMVNFYSL 586  
 | | | | |  
 QY 1 FAMPNFYTL 9

RESULT 6 PRELIMINARY; PRT; 1084 AA.  
 ID O64173  
 AC O64173;  
 DT 01-AUG-1998 (TrEMBLrel. 07, Created)  
 DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)  
 DT 01-MAY-1999 (TrEMBLrel. 10, Last annotation update)  
 DE RIBONUCLEOTIDE REDUCTASE LARGE SUBUNIT.  
 GN BRNDE.  
 OS Bacteriophage SPBc2.  
 OC Viruses; dsDNA viruses, no RNA stage; Tailed phages; Siphoviridae.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 98132653.  
 RA LAZAREVIC V., SOLDI B., DUESTERHOEFFT A., HILBERT H., MAUEL C.,  
 RA KARAMATA D.;  
 RT "Introns and intein coding sequence in the ribonucleotide reductase  
 RT genes of *Bacillus subtilis* temperate bacteriophage SPBeta.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 95:1692-1697(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA LAZAREVIC V., DUESTERHOEFFT A., SOLDI B., HILBERT H., MAUEL C.,  
 RA KARAMATA D.;  
 RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF020713; AAC13134.1; -;  
 DR PRAM; PF00317; ribonucleo-red; 2;  
 SQ SEQUENCE 1084 AA; 124620 MW; A212699F CRC32;

Query Match 73.3%; Score 55; DB 9; Length 1084;  
 Best Local Similarity 66.7%; Pred. No. 5.30e+00;  
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 876 FMVNFYSL 884  
 | | | | |  
 QY 1 FAMPNFYTL 9

RESULT 7 PRELIMINARY; PRT; 189 AA.  
 ID O26218  
 AC O26218;  
 DT 01-JAN-1998 (TrEMBLrel. 05, Created)  
 DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)  
 DT 01-AUG-1998 (TrEMBLrel. 07, Last annotation update)  
 DE HYPOTHETICAL 21.7 KD PROTEIN.  
 GN MTH115.  
 OS Methanobacterium thermoautotrophicum.  
 OC Archaea; Euryarchaeota; Methanobacteriales; Methanobacteriaceae;  
 OC Methanobacterium.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-DELTA H;  
 RX MEDLINE; 98037514.  
 RA SMITH D.R., DOUCETTE-STAMM L.A., DELOUGHERY C., LEE H.-M., DUBOIS J.,  
 RA ALDREDGE T., BASHIRZADEH R., BLAKELY D., COOK R., GILBERT K.,  
 RA HARRISON D., HOANG L., KEAGLE P., LUM W., POTIER B., QIU D.,  
 RA SPADAFORA R., VICARE R., WANG Y., WIERZBOWSKI J., GIBSON R.,  
 RA JIWANI N., CARUSO A., BUSH D., SAFER H., PATWELL D., PRABHAKAR S.,  
 RA MCDUGALL S., SHIMER G., GOYAL A., PIETROVSKI S., CHURCH G.M.,  
 RA DANIELS C.J., MAO J.-I., RICE P., NOLLING J., REEVE J.N.;  
 RT "Complete genome sequence of *Methanobacterium thermoautotrophicum*  
 RT *deltaH*: functional analysis and comparative genomics.";  
 RL J. Bacteriol. 179:7135-7155(1997).  
 DR EMBL; AE000801; AAB84621.1; -;  
 KW Hypothetical protein.  
 SQ SEQUENCE 189 AA; 21688 MW; D21257D8 CRC32;

Query Match 72.0%; Score 54; DB 1; Length 189;  
 Best Local Similarity 66.7%; Pred. No. 8.29e+00;  
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 100 FTMPSYTL 108  
 | | | | |  
 QY 1 FAMPNFYTL 9

RESULT 8 PRELIMINARY; PRT; 206 AA.  
 ID O59225  
 AC O59225;  
 DT 01-AUG-1998 (TrEMBLrel. 07, Created)  
 DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)  
 DT 01-JAN-1999 (TrEMBLrel. 09, Last annotation update)  
 DE 206AA LONG HYPOTHETICAL PROTEIN.  
 GN PH1547.  
 OS Pyrococcus horikoshii.  
 OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-OT3;  
 RX MEDLINE; 98344137.  
 RA KAWARABAYASHI Y., SAWADA M., HORIKAWA H., HAIKAWA Y., HINO Y.,  
 RA YAMAMOTO S., SEKINE M., BABU S., KOSUGI H., HOSOFYAMA A., NAGAI Y.,  
 RA SAKAI M., OGURA K., OTUKA R., NAKAZAWA H., TAKAMIYA M., OHFUKU Y.,  
 RA FUNAHASHI T., TANAKA T., KUDOH Y., YAMAZAKI J., KUSHIDA N., OSUCHI A.,  
 RA AKI K., NAKAMURA Y., ROBB T.F., HORIKOSHI K., MASUCHI Y., SHIZUFA H.,  
 RA KIKUCHI H.;  
 RT "Complete sequence and gene organization of the genome of a hyper-  
 RT thermophilic archaeobacterium, *Pyrococcus horikoshii* OT3.";  
 RL DNA Res. 5:55-76(1998).  
 DR EMBL; AP000006; BAA30659.1; -;  
 SQ SEQUENCE 206 AA; 24035 MW; BAA9E539 CRC32;

```

ID O48991 PRELIMINARY; PRT; 263 AA.
AC O48991;
DT 01-JUN-1998 (TREMELrel. 06, Created)
DT 01-JUN-1998 (TREMELrel. 06, Last sequence update)
DT 01-NOV-1999 (TREMELrel. 12, Last annotation update)
DE NBS-LRR TYPE RESISTANCE PROTEIN (FRAGMENT).
GN R11.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
OC Poaceae; Oryza.
RN [1].
RP SEQUENCE FROM N.A.
RC STRAIN=CV. NIPPONBARE;
RX MEDLINE; 9801880.
RA LEISTER D., KURTH J., LAURIE D.A., YANO M., SASAKI T., DEVOS K.,
RA GRANGER A., SCHULZE-LEFERT P.;
RT "Rapid reorganization of resistance gene homologues in cereal
RT genomes.";
RL Proc. Natl. Acad. Sci. U.S.A. 95:370-375(1998).
DR EMBL; AF032698; AAB96995.1; -.
DR MENDEL; 27306; Oryza;1426;27306.
DR PFAM; PF00931; NB-ARC; 1.
FT NON_TER 1
FT NON_TER 263
SQ SEQUENCE 263 AA; 30261 MW; 4C7583A9 CRC32;

Query Match 78.7%; Score 59; DB 10; Length 263;
Best Local Similarity 75.0%; Pred. No. 8.43e-01;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 118 TPNFYSL 125
QY 2 AMPNFYTL 9

RESULT 3
ID O29900 PRELIMINARY; PRT; 319 AA.
AC O29900;
DT 01-JAN-1998 (TREMELrel. 05, Created)
DT 01-JAN-1998 (TREMELrel. 05, Last sequence update)
DT 01-AUG-1998 (TREMELrel. 07, Last annotation update)
DE C4-DICARBOXYLATE TRANSPORTER (MAE1).
GN AF0347.
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobales; Archaeoglobaceae;
OC Archaeoglobus.
RN [1].
RP SEQUENCE FROM N.A.
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE; 98049343.
RA KLENK H.-P., CLAYTON R.A., TOMB J.-F., WHITE O., NELSON K.E.,
RA KETCHUM K.A., DODSON R.J., GWINN M., HICKY E.K., PETERSON J.D.,
RA RICHARDSON D.L., KERLAVAGE A.R., GRAHAM D.E., KYRDES N.C.,
RA FLEISCHMANN R.D., QUACKENBUSH J., LEE N.H., SUTTON G.G., GILL S.,
RA KIRKNESS E.F., DOUGHERTY B.A., MCKENNEY K., ADAMS M.D., LOFTUS B.,
RA PETERSON S., REICH C.I., MCNEIL L.K., BADGER J.H., GLODEK A., ZHOU L.,
RA OVERBECK R., GOCAYNE J.D., WEIDMAN J.F., McDONALD L., UTTERBACK T.,
RA COTTON M.D., SPRIGGS T., ARTIACH P., KAINE B.P., SYKES S.M.,
RA SADOW P.W., D'ANDREA K.P., BOWMAN C., FUJII C., GARLAND S.A.,
RA MASON T.M., OLSEN G.J., FRASER C.M., SMITH H.O., WOESE C.R.,
RA VENTER J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
RT reducing archaeon Archaeoglobus fulgidus.";
RL Nature 390:364-370(1997).
DR EMBL; AF001080; AAB90885.1; -.
DR TIGR; AF0347; -.
KW Hypothetical protein.
SQ SEQUENCE 319 AA; 35267 MW; 6F1B082 CRC32;

Query Match 74.7%; Score 56; DB 1; Length 319;
Best Local Similarity 66.7%; Pred. No. 3.37e+00;
Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

Db 68 FVGNFYPL 76
QY 1 FAMPNFYTL 9

RESULT 4
ID Q22625 PRELIMINARY; PRT; 418 AA.
AC Q22625;
DT 01-NOV-1996 (TREMELrel. 01, Created)
DT 01-NOV-1996 (TREMELrel. 01, Last sequence update)
DT 01-JAN-1999 (TREMELrel. 09, Last annotation update)
DE T21B10.4 PROTEIN.
GN T21B10.4.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
OC Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN [1].
RP SEQUENCE FROM N.A.
RC BAYNES C.;
RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
RN [2].
RP SEQUENCE FROM N.A.
RX MEDLINE; 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M., COULSON A.,
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., FULTON L.,
RA CRAYTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
DR EMBL; 268318; CAA92694.1; -.
SQ SEQUENCE 418 AA; 48023 MW; 8A0366AF CRC32;

Query Match 73.3%; Score 55; DB 5; Length 418;
Best Local Similarity 75.0%; Pred. No. 5.30e+00;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 295 SMPNFYEL 302
QY 2 AMPNFYTL 9

RESULT 5
ID O31874 PRELIMINARY; PRT; 786 AA.
AC O31874;
DT 01-JAN-1998 (TREMELrel. 05, Created)
DT 01-JAN-1998 (TREMELrel. 05, Last sequence update)
DT 01-MAY-1999 (TREMELrel. 10, Last annotation update)
DE YOSO PROTEIN (RIBONUCLEOTIDE REDUCTASE HOMOLOGY).
GN YOSO OR YOUP.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
RN [1].
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE; 98044033.
RA KUNST F., OGASAWARA N., MOSZER I., ALBERTINI A.M., ALLONI G.,
RA AZEVEDO V., BERTERO M.G., BESSIERES P., BOLOTIN A., BORCHERT S.,
RA BORRIS R., BOURSIER L., BRANS A., BRAUN M., BRIGHELLI S.C., BRON S.,
RA BROUILLET S., BRUSCHI C.V., CALDWELL B., CAPUANO V., CARTER N.M.,
RA CHOI S.K., CODANI J.J., CONNERTON I.F., CUMMINGS N.J., DANIEL R.A.,
RA DENIZOT F., DEVINE K.M., DUSTERHOF A., EHRLICH S.D., EMMERSON P.T.,
RA ENTIAN K.D., ERRINGTON J., FABRET C., FERRARI E., FOULGER D.,
RA FRITZ C., FUJITA M., FUJITA Y., FUMA S., GALIZZI A., GALLERON N.,
RA CHIM S.Y., GLASER P., GOFFEAU A., GOLIGHTLY E.J., GRANDI G.,

```

\*\*\*\*\*  
W P S R L  
(TM)  
\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Fri Apr 14 23:32:39 2000; MasPar time 11.13 Seconds  
Tabular output not generated. 56.066 Million cell updates/sec

Title: >US-08-452-843-6  
Description: (1-9) from US08452843.pap  
Perfect Score: 75  
Sequence: 1 FAMPNFYTL 9

Scoring table: PAM 150  
Gap 15

Searched: 225878 seqs, 69334122 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: sptrembl12  
1:sp\_archaea 2:sp\_bacteria 3:sp\_fungi 4:sp\_human  
5:sp\_invertebrate 6:sp\_mammal 7:sp\_mhc 8:sp\_organelle  
9:sp\_phase 10:sp\_plant 11:sp\_rodent 12:sp\_unclassified  
13:sp\_vertebrate 14:sp\_virus

Statistics: Mean 24.737; Variance 32.470; scale 0.762

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description	Pred. No.
1	60	80.0	329	2	Q9X0U9	OLIGOPEPTIDE ABC TRANS	5.26e-01
2	59	78.7	263	10	O48991	NBS-LRR TYPE RESISTANC	8.43e-01
3	56	74.7	319	1	O29900	C4-DICARBOXYLATE TRANS	3.37e+00
4	55	73.3	418	5	O22625	T21B10.4 PROTEIN.	5.30e+00
5	55	73.3	786	2	O31874	YOSO PROTEIN (RIBONUCLE	5.30e+00
6	55	73.3	1084	9	O64173	RIBONUCLEOTIDE REDUCTA	5.30e+00
7	54	72.0	189	1	O26218	HYPOTHETICAL 21.7 KD P	8.29e+00
8	54	72.0	206	1	O59225	206AA LONG HYPOTHETICA	8.29e+00
9	54	72.0	339	2	O05943	CYTOCHROME D UBIQUINOL	8.29e+00
10	54	72.0	466	5	O17451	GAG-LIKE PROTEIN.	8.29e+00
11	54	72.0	2763	13	O9XGV2	NEUROFIBROMATOSIS TYPE	8.29e+00
12	54	72.0	2764	5	O01399	NEUROFIBROMIN.	8.29e+00
13	54	72.0	2802	5	O01397	NEUROFIBROMIN.	8.29e+00
14	54	72.0	2802	5	O01398	NEUROFIBROMIN.	8.29e+00
15	54	72.0	2820	11	P97526	NEUROFIBROMIN.	8.29e+00
16	53	70.7	743	5	O23869	D2 ORF.	1.29e+01
17	53	70.7	763	5	O23874	D2 ORF.	1.29e+01
18	52	69.3	260	1	O59032	HYPOTHETICAL PROTEIN M	1.99e+01
19	52	69.3	262	5	O97429	MDG1HET PROTEIN (FRAGM	1.99e+01
20	52	69.3	263	2	O87983	ABC TRANSPORTER INTEGR	1.99e+01

21	52	69.3	275	5	017711	C55A1.1 PROTEIN.	1.99e+01
22	52	69.3	341	5	018101	T21B4.5 PROTEIN.	1.99e+01
23	52	69.3	364	5	P91384	COSMID K12D9.	1.99e+01
24	52	69.3	372	2	051368	MANNOSE-6-PHOSPHATE IS	1.99e+01
25	52	69.3	426	10	023842	S GLYCOPROTEIN (FRAGME	1.99e+01
26	52	69.3	429	10	023845	S GLYCOPROTEIN (FRAGME	1.99e+01
27	52	69.3	429	10	080346	S GLYCOPROTEIN (FRAGME	1.99e+01
28	52	69.3	931	14	P87544	104K PROTEIN.	1.99e+01
29	51	68.0	201	2	092D41	HYPOTHETICAL 23.3 KD P	3.06e+01
30	51	68.0	307	2	P95159	HYPOTHETICAL 33.2 KD P	3.06e+01
31	51	68.0	332	5	044531	K08A5.2 PROTEIN.	3.06e+01
32	51	68.0	605	2	P72607	ABC TRANSPORTER.	3.06e+01
33	51	68.0	706	3	O93884	DIHYDROXYACETONE SYNTH	3.06e+01
34	51	68.0	1464	14	O66951	E2 GLYCOPROTEIN PRECUR	3.06e+01
35	51	68.0	1798	5	O9XW15	Y54E2A.6 PROTEIN.	3.06e+01
36	50	66.7	103	1	O59400	103AA LONG HYPOTHETICA	4.67e+01
37	50	66.7	198	1	O59359	198AA LONG HYPOTHETICA	4.67e+01
38	50	66.7	199	11	O61907	PHOSPHATIDYLETHANOLAMI	4.67e+01
39	50	66.7	233	5	O77340	PFC0555C PROTEIN.	4.67e+01
40	50	66.7	306	5	O20457	F46C3.2 PROTEIN.	4.67e+01
41	50	66.7	355	5	O19572	CONTAINS SIMILARITY TO	4.67e+01
42	50	66.7	502	5	O21291	K07F5.6 PROTEIN.	4.67e+01
43	50	66.7	605	2	O84063	FLAGELLAR SECRETION PR	4.67e+01
44	50	66.7	980	5	O94251	K04A8.6 PROTEIN.	4.67e+01
45	50	66.7	1001	5	001261	T20D3.9 PROTEIN.	4.67e+01

ALIGNMENTS

RESULT	1	PRELIMINARY;	PRT;	329 AA.
ID	Q9X0U9			
AC	Q9X0U9			
DT	01-NOV-1999 (Tremblrel. 12, Created)			
DT	01-NOV-1999 (Tremblrel. 12, Last sequence update)			
DT	01-NOV-1999 (Tremblrel. 12, Last annotation update)			
DE	OLIGOPEPTIDE ABC TRANSPORTER, PERMEASE PROTEIN.			
GN	TM1222.			
OS	Thermotoga maritima.			
OC	Bacteria; Thermotogales; Thermotoga.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE; 99287316.			
RA	NELSON K.E., CLAYTON R.A., GILL S.R., GWINN M.L., DODSON R.J.,			
RA	HAFT D.H., HICKEY E.K., PETERSON J.D., NELSON W.C., KETCHUM K.A.,			
RA	MCDONALD L., UTTERBACK T.R., MALEK J.A., LINHER K.D., GARRETT M.M.,			
RA	STEWART A.M., COTTON M.D., PRATT M.S., PHILLIPS C.A., RICHARDSON D.,			
RA	HEIDELBERG J., SUTTON G.G., FLEISCHMANN R.D., WHITE O., SALZBERG S.L.,			
RA	SMITH H.O., VENTER J.C., FRASER C.M.;			
RT	"Evidence for lateral gene transfer between Archaea and bacteria from			
RT	genome sequence of Thermotoga maritima."			
RL	Nature 399:323-329(1999).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RA	NELSON K.E., CLAYTON R.A., GILL S.R., GWINN M.L., DODSON R.J.,			
RA	HAFT D.H., HICKEY E.K., PETERSON J.D., NELSON W.C., KETCHUM K.A.,			
RA	MCDONALD L., UTTERBACK T.R., MALEK J.A., LINHER K.D., GARRETT M.M.,			
RA	STEWART A.M., COTTON M.D., PRATT M.S., PHILLIPS C.A., RICHARDSON D.,			
RA	HEIDELBERG J., SUTTON G.G., FLEISCHMANN R.D., WHITE O., SALZBERG S.L.,			
RA	SMITH H.O., VENTER J.C., FRASER C.M.;			
RL	Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; AE001778; AAD36297.1; -			
SQ	SEQUENCE 329 AA; 37507 MW; B58267D2 CRC32;			

Query Match 80.0%; Score 60; DB 2; Length 329;  
Best Local Similarity 77.8%; Pred. No. 5.26e-01;  
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
Db 135 FALPFFYTL 143  
||| ||||  
QY 1 FAMPNFYTL 9

RESULT 2

THIS PAGE BLANK (USPTO)

RA COTTON M.D., WEIDMAN J.M., FUJII C., BOWMAN C., WATTHEY L., WALLIN E.,  
RA HAYES W.S., BORODOVSKY M., KARP P.D., SMITH H.O., FRASER C.M.,  
RA VENTER J.C.;  
RT "The complete genome sequence of the gastric pathogen Helicobacter  
pylori."  
RL Nature 388:539-547(1997).  
CC -!- SIMILARITY: BELONGS TO THE UPF0024 FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AE000602; AAD07971.1; -  
DR TIGR; HP0926;  
DR PROSITE; PS01268; UPF0024; 1.  
DR PFAM; PF01142; UPF0024; 1.  
KW Hypothetical protein.  
SQ SEQUENCE 381 AA; 44003 MW; C659A962 CRC32;

Query Match 66.7%; Score 50; DB 1; Length 381;  
Best Local Similarity 57.1%; Pred. No. 2.22e+01;  
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 159 FGMPNPF 165  
|:|:|:|:  
Qy 1 FAMPNPFY 7

Search completed: Fri Apr 14 23:32:19 2000  
Job time : 47 secs.



```
DR PIR: S01601; S01601.
DR MENDEL; 9017; MARPO:ndhj;1.
DR PROSITE; PS00542; COMPLEX1_30K; 1.
DR PFAM; PF00329; complex1_30Kd; 1.
KW Oxidoreductase; NAD; Plastoquinone; Chloroplast.
SQ SEQUENCE 169 AA; 20085 MW; 8BC10865 CRC32;

Query Match 66.7%; Score 50; DB 1; Length 169;
Best Local Similarity 55.6%; Pred. No. 2.22e+01;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 157 YIVPNEYEL 165
QY : : : : :
1 FAMPNEYTL 9

RESULT 13
ID PEWT_RAT STANDARD; PRT; 198 AA.
AC Q08388;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DE 01-OCT-1996 (Rel. 34, Last annotation update)
DE PHOSPHATIDYLETHANOLAMINE N-METHYLTRANSFERASE (EC 2.1.1.17) (PEMT).
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RX MEDLINE; 93346418.
RA CUI Z., VANCE J.E., CHEN M.H., VOELKER D.R., VANCE D.E.;
RT "Cloning and expression of a novel phosphatidylethanolamine N-
RT methyltransferase. A specific biochemical and cytological marker for
RT a unique membrane fraction in rat liver.";
RL J. Biol. Chem. 268:16655-16663(1993).
RN [2]
RP SEQUENCE OF 1-30.
RA RIDGWAY N.D.;
RA Thesis (1988), University of British Columbia, Canada.
CC -!- CATALYTIC ACTIVITY: S-ADENOSYL-L-METHIONINE + PHOSPHATIDYL-
CC ETHANOLAMINE -> S-ADENOSYL-L-HOMOCYSTEINE + PHOSPHATIDYL-N-
CC METHYLETHANOLAMINE.
CC -!- PATHWAY: FIRST, SECOND AND THIRD STEPS OF PHOSPHATIDYLETHANOLAMINE
CC METHYLATION PATHWAY.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -!- TISSUE SPECIFICITY: LIVER.
CC -!- SIMILARITY: TO YEAST PEM2.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L14441; AAA03154.1; -.
DR Phospholipid biosynthesis; Transferase; Methyltransferase;
KW Transmembrane.
FT INIT MET 0 0
FT TRANSMEM 12 32 POTENTIAL.
FT TRANSMEM 45 65 POTENTIAL.
FT TRANSMEM 90 110 POTENTIAL.
FT TRANSMEM 158 178 POTENTIAL.
SQ SEQUENCE 198 AA; 22355 MW; 90AE9A9B CRC32;

Query Match 66.7%; Score 50; DB 1; Length 198;
Best Local Similarity 44.4%; Pred. No. 2.22e+01;
Matches 4; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 102 FVLSFPYAL 110
QY : : : : :
1 FAMPNEYTL 9

RESULT 14
ID SPSP_BACSU STANDARD; PRT; 246 AA.
AC P39629;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DE 01-FEB-1996 (Rel. 33, Last annotation update)
DE SPORE COAT POLYSACCHARIDE BIOSYNTHESIS PROTEIN SPSP.
GN SPSP OR IPA-71D.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE; 95020537.
RA GLASER P., KUNST F., ARNAUD M., COUDART M.P., GONZALES W.,
RA HULLO M.F., IONESCU M., LUBOCHINSKY B., MARCELINO L., MOSZER I.,
RA PRESECAN E., SANTANA M., SCHNEIDER E., SCHWEIZER J., VERTES A.,
RA RAPOPORT G., DANCHIN A.;
RT "Bacillus subtilis genome project: cloning and sequencing of the 97
RT kb region from 325 degrees to 333 degrees.";
RL Mol. Microbiol. 10:371-384(1993).
CC -!- PATHWAY: SPORE COAT POLYSACCHARIDE BIOSYNTHESIS.
CC -!- SIMILARITY: BELONGS TO THE GLUCOSE-1-PHOSPHATE
CC THYMIDYLTRANSFERASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X73124; CAAS1627.1; -.
DR EMBL; X79123; CAB13810.1; -.
DR SUBTILIST; BG10617; SPSP.
DR PFAM; PF00483; NTP transferase; 1.
KW Transferase; Kinase; Nucleotidyltransferase.
SQ SEQUENCE 246 AA; 27773 MW; 921EF443 CRC32;

Query Match 66.7%; Score 50; DB 1; Length 246;
Best Local Similarity 71.4%; Pred. No. 2.22e+01;
Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 58 MPQFYKL 64
QY : : : : :
3 MPNFYTL 9

RESULT 15
ID Y926_HELPY STANDARD; PRT; 381 AA.
AC P53985;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE 01-NOV-1997 (Rel. 35, Last annotation update)
DE HYPOTHETICAL PROTEIN HP0926.
GN HP0926
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=26695 / ATCC 700392;
RX MEDLINE; 97394467.
RA TOMB J.-F., WHITE O., KERLAVAGE A.R., CLAYTON R.A., SUTTON G.G.,
RA FLISCHMANN R.D., KETCHUM K.A., KLENK H.-P., GILL S., DOUGHERTY B.A.,
RA NELSON K., QUACKENBUSH J., ZHOU L., KIRKNESS E.F., PETERSON S.,
RA LOFTUS B., RICHARDSON D., DODSON R., KHALAK H.G., GLODEK A.,
RA MCKENNEY K., FITZGERALD L.M., LEE N., ADAMS M.D., HICKEY E.K.,
RA BERG D.E., GOCAYNE J.D., UTTERBACK T.R., PETERSON J.D., KELLEY J.M.,
```



```
FT DOMAIN 1237 1453 RAS-GAP.
FT VARSPLIC 1373 1393 MISSING (IN ISOFORM I AND ISOFORM IV).
FT VARSPLIC 1394 1406 VVSQFPQNSIGA -> VPKSCFCSLNNRWLASLRT
FT VARSPLIC 1407 2841 ASVP (IN ISOFORM III AND ISOFORM IV).
FT SEQUENCE 2841 AA: 319591 MW; A7AA76F4 CRC32;
Query Match 72.0%; Score 54; DB 1; Length 2841;
Best Local Similarity 55.6%; Pred. No. 4.06e+00;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
Db 2158 FSLPKFYLL 2166
|:|:|:|
QY 1 FAMPNFYTL 9

RESULT 8
ID YN8S YEAST STANDARD; PRT; 393 AA.
AC P53740;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DE 01-OCT-1996 (Rel. 34, Last annotation update)
DE HYPOHETICAL 44.5 KD PROTEIN IN PER494-MS01 INTERGENIC REGION.
GN YNR048W OR N3453.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
OC Saccharomycetaceae; Saccharomycetes.
RN [1]
RP SEQUENCE FROM N.A.
RA POHL T.M.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: TO YEAST YCR94W AND YNL323W.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: 271663; CAA96329.1;
KW Hypothetical protein; Transmembrane.
FT TRANSMEM 47 67 POTENTIAL.
FT TRANSMEM 335 355
SQ SEQUENCE 393 AA; 44542 MW; 4660346A CRC32;

Query Match 69.3%; Score 52; DB 1; Length 393;
Best Local Similarity 75.0%; Pred. No. 9.61e+00;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
Db 280 ALPNFYKL 287
|:|:|:|
QY 2 AMPNFYTL 9

RESULT 9
ID Y506_RICPR STANDARD; PRT; 201 AA.
AC Q9ZD41;
DT 15-DEC-1999 (Rel. 39, Created)
DT 15-DEC-1999 (Rel. 39, Last sequence update)
DT 15-DEC-1999 (Rel. 39, Last annotation update)
DE HYPOHETICAL PROTEIN RP506.
GN RP506.
OS Rickettsia prowazekii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MADRID E;
RX MEDLINE: 99039499.
RA ANDERSSON S.G.E., ZOMORODIPOUR A., ANDERSSON J.O.,
RA SICHERITZ-PONTEN T., ALSMARK U.C.M., PODOWSKI R.M., NAEGLUND A.K.,

RA ERIKSSON A.-S., WINKLER H.H., KURLAND C.G.;
RT "The genome sequence of Rickettsia prowazekii and the origin of
RL mitochondria.";
RL Nature 396:133-140(1998).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AJ235272; CAA14958.1;
KW Hypothetical protein; Transmembrane.
FT TRANSMEM 11 31 POTENTIAL.
FT SEQUENCE 201 AA; 23278 MW; 6B1D3C55 CRC32;

Query Match 68.0%; Score 51; DB 1; Length 201;
Best Local Similarity 71.4%; Pred. No. 1.47e+01;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
Db 1 MPSEFYKL 7
|:|:|:|
QY 3 MPNFYTL 9

RESULT 10
ID TYRA_LACLA STANDARD; PRT; 354 AA.
AC P43901;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE PREPHENATE DEHYDROGENASE (EC 1.3.1.12) (PDH).
GN TYRA.
OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Lactococcus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=F15876;
RX MEDLINE: 95124293
RA GRIFFIN H.G., GASSON M.J.;
RT "Genetic aspects of aromatic amino acid biosynthesis in Lactococcus
RT lactis.";
RL Mol. Gen. Genet. 246:119-127(1995).
CC -1- CATALYTIC ACTIVITY: PREPHENATE + NAD(+) -> 4-HYDROXYPHENYLPIRUVATE
CC + CO(2) + NADH.
CC -1- PATHWAY: TYROSINE BIOSYNTHESIS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X78413; CAA55179.1;
KW Tyrosine biosynthesis; Oxidoreductase; NAD.
FT AP_BIND 3 33530 MW; DD0FE758 CRC32;
FT SEQUENCE 354 AA; 39530 MW;
SQ SEQUENCE 354 AA; 39530 MW; DD0FE758 CRC32;

Query Match 68.0%; Score 51; DB 1; Length 354;
Best Local Similarity 75.0%; Pred. No. 1.47e+01;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
Db 281 ALPNFYDL 288
|:|:|:|
QY 2 AMPNFYTL 9

RESULT 11
```

RL Hum. Mol. Genet. 2:1861-1864(1993).  
RN [17]  
RX VARIANT NF1 ASN-2387--PHE-2388 DEL.  
RX MEDLINE: 94362704.  
RA ABERNATHY C.R., COLMAN S.D., KOUSSEFF B.G., WALLACE M.R.;  
RT "Two NF1 mutations: frameshift in the GAP-related domain, and loss of  
RT two codons toward the 3' end of the gene.";  
RN Hum. Mutat. 3:347-352(1994).  
RN [18]  
RX VARIANT NF1 ALA-2631.  
RX MEDLINE: 96091873.  
RA UPADHYAYA M., MAYNARD J., OSBORN M., HUSON S.M., PONDER M.,  
RA PONDER B.A.J., HARPER P.S.;  
RT "Characterisation of germline mutations in the neurofibromatosis type  
RT 1 (NF1) gene.";  
RN J. Med. Genet. 32:706-710(1995).  
RN [19]  
RX VARIANT NF1 ARG-629.  
RX MEDLINE: 96431167.  
RA GASPARINI P., D'AGROMA L., DE CILLIS G.P., BALESTRAZZI P.,  
RA MINGARELLI R., ZELANTE L.;  
RT "Scanning the first part of the neurofibromatosis type 1 gene by RNA-  
RT SSCP: identification of three novel mutations and of two new  
RT polymorphisms.";  
RN Hum. Genet. 97:492-495(1996).  
RN [20]  
RX VARIANT LS ARG-1035.  
RX MEDLINE: 96400960.  
RA WU R., LEGIUS E., ROBBERECHT W., DUMOULIN M., CASSIMAN J.-J.,  
RA FRYS J.-P.;  
RT "Neurofibromatosis type I gene mutation in a patient with features of  
RT LEOPARD syndrome.";  
RN Hum. Mutat. 8:51-56(1996).  
RN [21]  
RX VARIANTS NF1 ARG-844 AND PRO-898.  
RX MEDLINE: 97295087.  
RA MAYNARD J., KRAWCZAK M., UPADHYAYA M.;  
RT "Characterization and significance of nine novel mutations in exon 16  
RT of the neurofibromatosis type 1 (NF1) gene.";  
RN Hum. Genet. 99:674-676(1997).  
RN [22]  
RX VARIANT NF1 ARG-1952.  
RX MEDLINE: 97255969.  
RA HUDSON J., WU C.L., TASSABEHJI M., SUMMERS E.M., SIMON S., SUPER M.,  
RA DONNAI D., THAKKER N.;  
RT "Novel and recurrent mutations in the neurofibromatosis type 1 (NF1)  
RT gene.";  
RN Hum. Mutat. 9:366-367(1997).  
RN [23]  
RX VARIANT NF1 TRP-1611.  
RX MEDLINE: 97442280.  
RA UPADHYAYA M., MAYNARD J., OSBORN M., HARPER P.S.;  
RT "Six novel mutations in the neurofibromatosis type 1 (NF1) gene.";  
RN Hum. Mutat. 10:248-250(1997).  
CC -!- FUNCTION: STIMULATES THE GTPASE ACTIVITY OF RAS. NF1 SHOWS GREATER  
CC AFFINITY FOR RAS GAP, BUT LOWER SPECIFIC ACTIVITY. THUS IT MAY BE  
CC A REGULATOR OF RAS ACTIVITY.  
CC -!- ALTERNATIVE PRODUCTS: TWO ISOFORMS; I AND II (SHOWN HERE); ARE  
CC PRODUCED BY ALTERNATIVE SPLICING.  
CC -!- DISEASE: THIS PROTEIN IS ASSOCIATED WITH TYPE 1 NEUROFIBROMATOSIS  
CC (NF1) (ALSO CALLED VON RECKLINGHAUSEN SYNDROME), THE MOST FREQUENT  
CC INHERITED GENETIC DISEASE (ABOUT 1 IN 3000). IT EXHIBITS FULL  
CC PENETRANCE AND HIGH MUTATION RATE WITH 30 TO 50% OF NF1 PATIENTS  
CC REPRESENTING A NEW MUTATION. AMONG THE MANY CLINICAL FEATURES OF  
CC NF1 ARE PATCHES OF SKIN PIGMENTATION (CAFE-AU-LAIT SPOTS), LISH  
CC NODULES OF THE IRIS PERIPHERAL, PERIPHERAL NERVOUS SYSTEM  
CC ASSOCIATED TUMORS AND FIBROMATOUS SKIN TUMORS. THE DISEASE  
CC DEMONSTRATES A HIGH DEGREE OF PENETRANCE BY AGE 5 YEARS.  
CC -!- DISEASE: DEFECTS IN NF1 ARE ASSOCIATED WITH WATSON SYNDROME (WS).  
...  
Note: remainder of annotations omitted.

Query Match 72.0%; Score 54; DB 1; Length 2839;

Best Local Similarity 55.6%; Pred. No. 4.06e+00;  
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;  
Db 2156 FSLPKFYLL 2164  
Qy 1 FAMPNFYLL 9  
I:|:|:|:|  
|:|:|:|:|  
RESULT 7  
ID NF1\_MOUSE STANDARD; PRT; 2841 AA.  
AC 004690; 061956; 061957;  
DT 01-JUN-1994 (Rel. 29, Created)  
DT 01-JUN-1994 (Rel. 29, Last sequence update)  
DT 15-DEC-1999 (Rel. 39, Last annotation update)  
DE NEUROFIBROMIN (NEUROFIBROMATOSIS-RELATED PROTEIN NF-1).  
GN NF1.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BALB/C; TISSUE-BRAIN;  
RX MEDLINE: 93357730.  
RA BERNARDS A., SNIJDERS A.J., HANNIGAN G.E., MURPHY A.E., GUSELLA J.F.;  
RT "Mouse neurofibromatosis type 1 cDNA sequence reveals high degree of  
RT conservation of both coding and non-coding mRNA segments.";  
RN Hum. Mol. Genet. 2:645-650(1993).  
RN [2]  
RP SEQUENCE OF 1178-1555 FROM N.A., AND ALTERNATIVE SPLICING.  
RX MEDLINE: 95047432.  
RA MANTANI A., MAKASUGI S., YOKOTA Y., ABE K., USHIO Y., YAMAMURA K.;  
RT "A novel isoform of the neurofibromatosis type-1 mRNA and a switch of  
RT isoforms during murine cell differentiation and proliferation.";  
RN Gene 148:245-251(1994).  
RN [3]  
RP SEQUENCE OF 1950-2568 FROM N.A.  
RX MEDLINE: 90384569.  
RA BUCHBERG A.M., CLEVELAND L.S., JENKINS N.A., COPELAND N.G.;  
RT "Sequence homology shared by neurofibromatosis type-1 gene and IRA-1  
RT and IRA-2 negative regulators of the Ras cyclic AMP pathway.";  
RN Nature 347:291-294(1990).  
CC -!- FUNCTION: STIMULATES THE GTPASE ACTIVITY OF RAS. NF1 SHOWS GREATER  
CC AFFINITY FOR RAS GAP, BUT LOWER SPECIFIC ACTIVITY. THUS IT MAY BE  
CC A REGULATOR OF RAS ACTIVITY.  
CC -!- ALTERNATIVE PRODUCTS: FOUR ISOFORMS; I, II (SHOWN HERE), III AND  
CC IV; ARE PRODUCED BY ALTERNATIVE SPLICING.  
CC -!- TISSUE SPECIFICITY: TYPE I IS EXPRESSED PREDOMINANTLY IN BRAIN/  
CC SPINAL CORD AND TESTIS. TYPE II IS EXPRESSED PREDOMINANTLY IN  
CC ADRENAL GLAND, KIDNEY, OVARY AND LUNG. TYPE III IS EXPRESSED  
CC PREDOMINANTLY IN ADRENAL GLAND AND TYPE IV IS EXPRESSED  
CC MAINLY IN THE TESTIS.  
CC -!- SIMILARITY: TO OTHER RAS GTPASE-ACTIVATING PROTEINS.  
CC -----  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; L10369; AAA39806.1; -  
CC EMBL; L10367; AAA39806.1; JOINED.  
CC EMBL; L10368; AAA39806.1; JOINED.  
CC EMBL; L10370; AAA68132.1; -  
CC EMBL; X54924; CAA38690.1; -  
CC EMBL; D30730; BAA06395.1; -  
CC EMBL; D30731; BAA06396.1; -  
CC MGD; MGI:97306; NF1.  
CC PROSITE; PS00509; RAS\_GTPASE\_ACTIV\_1; 1.  
CC PROSITE; PS50018; RAS\_GTPASE\_ACTIV\_2; 1.  
CC PFAM; PF00616; RASGAP; 1.  
KW Gtpase activation; Alternative splicing.

FT CARBOHYD 398 398 POTENTIAL.  
FT CONFLICT 397 H -> E (IN REF. 2).  
FT CONFLICT 567 N -> D (IN REF. 2).  
SQ SEQUENCE 753 AA; 86329 MW; 37CE9258 CRC32;

Query Match 72.08; Score 54; DB 1; Length 753;  
Best Local Similarity 55.68; Pred.No. 4.06e+00;  
Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

DB 641 FVMAAFYPL 649  
I I I I I  
QY 1 FAMPNEYTL 9

RESULT 6  
ID NF1\_HUMAN STANDARD; PRT; 2839 AA.  
AC P21359;  
DT 01-MAY-1991 (Rel. 18, Created)  
DT 01-JUN-1994 (Rel. 29, Last sequence update)  
DT 15-DEC-1999 (Rel. 39, Last annotation update)  
DE NEUROFIBROMIN (NEUROFIBROMATOSIS-RELATED PROTEIN NF-1).  
GN NF1.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.  
RN [1]  
RP SEQUENCE OF 1-1370 AND 1392-2839 FROM N.A.  
RX MEDLINE; 92147138.  
RA MARCHUK D.A., SAULINO A., TAVAKOL R., SWAROOP M., WALLACE M.R.,  
RA ANDERSEN L.B., MITCHELL A.L., GUTMANN D.H., BOGUSKI M., COLLINS F.S.;  
RT "CDNA cloning of the type 1 neurofibromatosis gene: complete sequence  
of the NF1 gene product.";  
RL Genomics 11:931-940(1991).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 93090270.  
RA BERNARDS A., HAASE V.H., MURPHY A.E., MENON A., HANNIGAN G.E.,  
RA GUSELJA J.F.;  
RT "Complete human NF1 cDNA sequence: two alternatively spliced mRNAs  
and absence of expression in a neuroblastoma line.";  
RL DNA Cell Biol. 11:727-734(1992).  
RN [3]  
RP SEQUENCE OF 335-1370 AND 1392-2839 FROM N.A.  
RX MEDLINE; 9035969.  
RA XU G., O'CONNELL P., VISKOCHIL D., CANTHON R., ROBERTSON M.,  
RA CULVER M., DUNN D., STEVENS J., GESTELAND R., WHITE R., WEISS R.;  
RT "The neurofibromatosis type 1 gene encodes a protein related to GAP.";  
RL Cell 62:599-608(1990).  
RN [4]  
RP SEQUENCE OF 1096-1370 AND 1372-1590 FROM N.A.  
RX MEDLINE; 91029515.  
RA MARTIN G.A., VISKOCHIL D., BOLLAG G., MCCABE P.C., CROSIER W.J.,  
RA HAUBRUCK H., CONROY L., CLARK R., O'CONNELL P., CANTHON R.M.,  
RA INNIS M., MCCORMICK F.;  
RT "The GAP-related domain of the neurofibromatosis type 1 gene product  
interacts with ras p21.";  
RL Cell 63:843-849(1990).  
RN [5]  
RP SEQUENCE OF 1606-2709 FROM N.A., AND VARIANT PRO-1953.  
RX MEDLINE; 90304909.  
RA CANTHON R.M., WEISS R., XU G., VISKOCHIL D., CULVER M., STEVENS J.,  
RA ROBERTSON M., DUNN D., GESTELAND R., O'CONNELL P., WHITE R.;  
RT "A major segment of the neurofibromatosis type 1 gene: cDNA sequence,  
genomic structure, and point mutations.";  
RL Cell 62:193-201(1990).  
RN [6]  
RP SEQUENCE OF 2230-2839 FROM N.A.  
RX MEDLINE; 90319792.  
RA WALLACE M.R., MARCHUK D.A., ANDERSEN L.B., LETCHER R., ODEH H.M.,  
RA SAULINO A.M., FOUNTAIN J.W., BRERETON A., NICHOLSON J., MITCHELL A.L.,  
RA BROWNSTEIN B.H., COLLINS F.S.;  
RT "Type 1 neurofibromatosis gene: identification of a large transcript  
disrupted in three NF1 patients.";

Science 249:181-186(1990).  
[7]  
RN ERRATUM.  
RX MEDLINE; 91102559.  
RA WALLACE M.R., MARCHUK D.A., ANDERSEN L.B., COLLINS F.S.;  
RL Science 250:1749-1749(1990).  
RN [8]  
RP SEQUENCE OF 1168-1566 FROM N.A.  
RX MEDLINE; 92019823.  
RA NISHI T., LEE P.S., OKA K., LEVIN V.A., TANASE S., MORINO Y.,  
RA SAYA H.;  
RT "Differential expression of two types of the neurofibromatosis type 1  
(NF1) gene transcripts related to neuronal differentiation.";  
RL Oncogene 6:1555-1559(1991).  
RN [9]  
RP SEQUENCE OF 1371-1391 FROM N.A.  
RX MEDLINE; 93109335.  
RA ANDERSEN L.B., BALLESTER R., MARCHUK D.A., CHANG E., GUTMANN D.H.,  
RA SAULINO A.M., CAMONIS J., WIGLER M., COLLINS F.S.;  
RT "A conserved alternative splice in the von Recklinghausen  
neurofibromatosis (NF1) gene produces two neurofibromin isoforms,  
both of which have GTPase-activating protein activity.";  
RL Mol. Cell. Biol. 13:487-495(1993).  
RN [10]  
RP FUNCTION.  
RX MEDLINE; 91029516.  
RA BALLESTER R., MARCHUK D., BOGUSKI M.S., SAULINO A., LETCHER R.,  
RA WIGLER M., COLLINS F.S.;  
RT "The NF1 locus encodes a protein functionally related to mammalian  
GAP and yeast IRA proteins.";  
RL Cell 63:851-859(1990).  
RN [11]  
RP REVIEW ON VARIANTS.  
RX MEDLINE; 95072625.  
RA UPADHYAYA M., SHAW D.J., HARPER P.S.;  
RT "Molecular basis of neurofibromatosis type 1 (NF1): mutation analysis  
and polymorphisms in the NF1 gene.";  
RL Hum. Mutat. 4:83-101(1994).  
RN [12]  
RP REVIEW ON VARIANTS.  
RX MEDLINE; 96422425.  
RA HONG SHEN M., HARPER P.S., UPADHYAYA M.;  
RT "Molecular genetics of neurofibromatosis type 1 (NF1).";  
RL J. Med. Genet. 33:2-17(1996).  
RN [13]  
RP VARIANT GLU-1444.  
RX MEDLINE; 92233464.  
RA LI Y., BOLLAG G., CLARK R., STEVENS J., CONROY L., FULTS D., WARD K.,  
RA FRIEDMAN E., SAMOWITZ W., ROBERTSON M., BRADLEY P., MCCORMICK F.,  
RA WHITE R., CANTHON R.;  
RT "Somatic mutations in the neurofibromatosis 1 gene in human tumors.";  
RL Cell 69:275-281(1992).  
RN [14]  
RP VARIANTS MET-2164 AND ASN-2192.  
RX MEDLINE; 93258316.  
RA UPADHYAYA M., SHEN M., CHERRYSON A., FARNHAM J., MAYNARD J.,  
RA HUSON S.M., HARPER P.S.;  
RT "Analysis of mutations at the neurofibromatosis 1 (NF1) locus.";  
RL Hum. Mol. Genet. 1:735-740(1992).  
RN [15]  
RP VARIANT HTS-1721--LEU-1733 DUPLICATION.  
RX MEDLINE; 93304433.  
RA TASSABEHI M., SPRACHAN T., SHARLAND M., COLLEY A., DONNAI D.,  
RA HARRIS R., THAKKER N.;  
RT "Tandem duplication within a neurofibromatosis type 1 (NF1) gene exon  
in a family with features of Watson syndrome and Noonan syndrome.";  
RL Am. J. Hum. Genet. 53:90-95(1993).  
RN [16]  
RP VARIANT MET-991 DEL.  
RX MEDLINE; 94108439.  
RA SHEN M.H., HARPER P.S., UPADHYAYA M.;  
RT "Neurofibromatosis type 1 (NF1): the search for mutations by PCR-  
heteroduplex analysis on Hydrolink gels.";

Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 31 FSLPKFYLL 39

QY 1 FAMPNFYLL 9

## RESULT 4

ID POP2 YEAST STANDARD; PRT; 433 AA.  
AC P39008; 1995 (Rel. 31, Created)  
DT 01-FEB-1996 (Rel. 34, Last sequence update)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 15-JUL-1998 (Rel. 36, Last annotation update)  
DE POP2 PROTEIN (CCR4-ASSOCIATED FACTOR 1).  
GN POP2 OR CAF1 OR YNR052C OR N3470  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;  
OC Saccharomycetaceae; Saccharomycetes.  
[1]  
RN SEQUENCE FROM N.A.  
RC STRAIN-S288C, AND A364A;  
RX MEDLINE; 93117094  
RA SAKAI A., CHIBAZAKURA T., SHIMIZU Y., HISHINUMA F.;  
RT "Molecular analysis of POP2 gene, a gene required for glucose-  
RT derepression of gene expression in Saccharomyces cerevisiae.";  
RT Nucleic Acids Res. 20:6227-6233(1992).  
[2]  
RN SEQUENCE FROM N.A.  
RA FOHL T.M.;  
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.  
[3]  
RN SEQUENCE OF 213-433 FROM N.A.  
RA CUSICK M.E.;  
RL Submitted (MAR-1992) to the EMBL/GenBank/DBJ databases.  
[4]  
RN CHARACTERIZATION.  
RX MEDLINE; 95311945.  
RA DRAPER M.P., SALVADORE C., DENIS C.L.;  
RT "Identification of a mouse protein whose homolog in Saccharomyces  
RT cerevisiae is a component of the CCR4 transcriptional regulatory  
RT complex.";  
RL Mol. Cell. Biol. 15:3487-3495(1995).  
CC -1- FUNCTION: UBIQUITOUS TRANSCRIPTION FACTOR REQUIRED FOR A DIVERSE  
CC SET OF PROCESSES. IT IS A COMPONENT OF THE CCR4 COMPLEX INVOLVED  
CC IN THE CONTROL OF ADH2 GENE EXPRESSION.  
CC -1- SIMILARITY: TO VERTEBRATE AND C.ELEGANS CAF1.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; D12807; BAA02246.1; -  
CC EMBL; D12808; BAA02247.1; -  
CC EMBL; D71667; CAA96333.1; -  
CC EMBL; M88607; AAA34832.1; -  
CC PIR; S35997; S35997.  
CC PIR; S35996; S35996.  
CC SGD; L0001465; POP2.  
KW Transcription regulation; Repressor.  
FT DOMAIN 81 90  
FT POLY-GLN.  
FT DOMAIN 111 125  
FT POLY-GLN.  
FT DOMAIN 363 369  
FT POLY-GLN.  
FT VARIANT 41 41  
FT K -> Q (IN STRAIN A364A).  
FT VARIANT 91 91  
FT Q -> QQQQQQQQQQQQQQQQ (IN STRAIN  
FT A364A).  
FT VARIANT 118 122  
FT MISSING (IN STRAIN A364A).  
FT VARIANT 278 278  
FT L -> S (IN STRAIN A364A).  
FT VARIANT 412 412  
FT K -> M.  
SQ SEQUENCE 433 AA; 49682 MW; E8582846 CRC32;

Query Match 72.0%; Score 54; DB 1; Length 433;  
Best Local Similarity 85.7%; Pred. No. 4.06e+00;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 338 MPNFYDL 344

QY 3 MPNFYTL 9

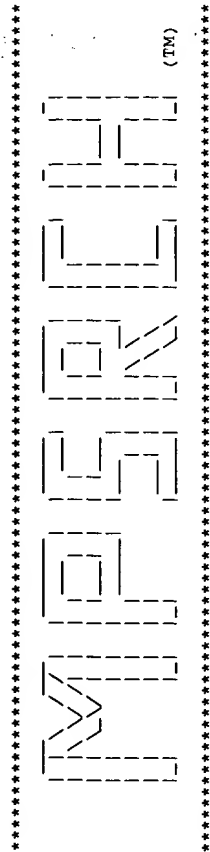
## RESULT 5

ID PMT3 YEAST STANDARD; PRT; 753 AA.  
AC P47190;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE DOLICHYL-PHOSPHATE-MANNOSE--PROTEIN MANNOsylTRANSFERASE 3  
DE (EC 2.4.1.109).  
GN PMT3 OR YOR321W OR O6148.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;  
OC Saccharomycetaceae; Saccharomycetes.  
[1]  
RN SEQUENCE FROM N.A.  
RX MEDLINE; 96158055.  
RA IMMERVOLL T., GENTZSCH M., TANNER W.;  
RT "PMT3 and PMT4, two new members of the protein-O-mannosyltransferase  
RT gene family of Saccharomyces cerevisiae.";  
RL Yeast 11:1345-1351(1995).  
[2]  
RN SEQUENCE FROM N.A.  
RC STRAIN-S288C / FY1679;  
RX MEDLINE; 97051589.  
RA PEARSON B.M., HERNANDO Y., PAYNE J., WOLF S.S., KALOGEROPOULOS A.,  
RA SCHWEIZER M.;  
RT "Sequencing of a 35.71 kb DNA segment on the right arm of yeast  
RT chromosome XV reveals regions of similarity to Chromosomes I and  
RT XIII.";  
RL Yeast 12:1021-1031(1996).  
CC -1- FUNCTION: TRANSFERS MANNOSE FROM DOL-P-MANNOSE TO SER OR THR  
CC RESIDUES ON PROTEINS (BY SIMILARITY).  
CC -1- CATALYTIC ACTIVITY: DOLICHYL PHOSPHATE D-MANNOSE + PROTEIN -  
CC DOLICHYL PHOSPHATE + O-D-MANNOsyl-PROTEIN.  
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. ENDOPLASMIC  
CC RETICULUM (BY SIMILARITY).  
CC -1- SIMILARITY: BELONGS TO THE FUNGAL PMT FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; X83797; CAA58728.1; -  
CC EMBL; X90565; CAA62176.1; -  
CC EMBL; Z75229; CAA99641.1; -  
CC SGD; L0002622; PMT3.  
KW Transferase; Glycosyltransferase; Glycoprotein; Transmembrane;  
KW Endoplasmic reticulum; Multigene family.  
FT TRANSMEM 51 71  
FT POTENTIAL.  
FT TRANSMEM 149 169  
FT POTENTIAL.  
FT TRANSMEM 175 195  
FT POTENTIAL.  
FT TRANSMEM 236 256  
FT POTENTIAL.  
FT TRANSMEM 283 303  
FT POTENTIAL.  
FT TRANSMEM 603 623  
FT POTENTIAL.  
FT TRANSMEM 640 660  
FT POTENTIAL.  
FT TRANSMEM 666 686  
FT POTENTIAL.  
FT TRANSMEM 704 724  
FT POTENTIAL.  
FT CARBOHYD 48 48  
FT POTENTIAL.  
FT CARBOHYD 124 124  
FT POTENTIAL.  
FT CARBOHYD 324 324  
FT POTENTIAL.

DT 01-FEB-1995 (Rel. 31, Last sequence update)  
DT 01-FEB-1995 (Rel. 31, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE MEMBRANE-BOUND LYTIC MUREIN TRANSGLYCOSYLASE B PRECURSOR (EC 3.2.1.1-)  
DE (MUREIN HYDROLASE B) (35 KD SOLUBLE LYTIC TRANSGLYCOSYLASE) (SLT35).  
GN MTB.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Escherichia.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-K12.  
RX MEDLINE; 96065704.  
RA EHLERT K., HOELTJE J.-V., TEMPLIN M.F.;  
RT "Cloning and expression of a murein hydrolase lipoprotein from  
RT Escherichia coli.";  
RL Mol. Microbiol. 16:761-768(1995).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 95309413.  
RA DIJKSTRA A.J., HERMANN F., KECK W.;  
RT "Cloning and controlled overexpression of the gene encoding the 35  
RT kDa soluble lytic transglycosylase from Escherichia coli.";  
RN FEBS Lett. 366:115-118(1995).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-K12 / MG1655;  
RX MEDLINE; 97426617.  
RA BLATTNER F.R., PLUNKETT G. III, BLOCH C.A., PERNA N.T., BURLAND V.,  
RA RILEY M., COLLADO-VIDES J., GLASNER F.D., RODE C.K., MAYHEW G.F.,  
RA GREGOR J., DAVIS N.W., KIRKPATRICK H.A., GOEDEN M.A., ROSE D.J.,  
RA MAU B., SHAO Y.;  
RT "The complete genome sequence of Escherichia coli K-12.";  
RN Science 277:1453-1474(1997).  
RN [4]  
RP SEQUENCE FROM N.A.  
RX STRAIN-K12;  
RX MEDLINE; 97349980.  
RA YAMATO Y., AIBA H., BABA T., HAYASHI K., INADA T., ISONO K.,  
RA ITOH T., KIMURA S., KITAGAWA M., MAKINO K., MIKI T., MITSUHASHI N.,  
RA MIZOBUCHI K., MORI H., NAKADE S., NAKAMURA Y., NASHIMOTO H.,  
RA OSHIMA T., OYAMA S., SAITO N., SAMPEI G., SATOH Y., SIVASUNDARAM S.,  
RA TAGAMI H., TAKAHASHI H., TAKEDA J., TAKEMOTO K., UEHARA K., WADA C.,  
RA YAMAGATA S., HORIUCHI T.;  
RT "Construction of a contiguous 874-kb sequence of the Escherichia coli  
RT - K12 genome corresponding to 50.0-68.8 min on the linkage map and  
RT analysis of its sequence features.";  
RN DNA Res. 4:91-113(1997).  
RN [5]  
RP PRELIMINARY SEQUENCE OF 1-91 FROM N.A.  
RX MEDLINE; 87194727.  
RA YAMADA M., SAIER M.H. JR.;  
RT "Glucitol-specific enzymes of the phosphotransferase system in  
RT Escherichia coli. Nucleotide sequence of the gut operon.";  
RN J. Biol. Chem. 262:5455-5463(1987).  
RN [6]  
RP X-RAY CRYSTALLOGRAPHY (1.7 ANGSTROMS) OF 42-361.  
RX MEDLINE; 98437484.  
RA VAN ASSELT E.J., PERRAKIS A., KALK K.H., LAMZIN V.S., DIJKSTRA B.W.;  
RT "Accelerated x-ray structure elucidation of a 36 kDa  
RT mureinase/transglycosylase using wARP.";  
RL Acta Crystallogr. D 54:58-73(1998).  
CC [1]- FUNCTION: MUREIN-DEGRADING ENZYME. MAY PLAY A ROLE IN RECYCLING  
CC OF MUROPETIDES DURING CELL ELONGATION AND/OR CELL DIVISION.  
CC [1]- CATALYTIC ACTIVITY: CLEAVAGE OF THE BETA-1,4-GLYCOSIDIC BOND  
CC BETWEEN N-ACETYLGLUCOSAMINE ACID AND N-ACETYLGLUCOSAMINE RESIDUES,  
CC THEREBY CONSERVING THE ENERGY IN A NEWLY SYNTHESIZED  
CC 1,6-ANHYDROBOND IN THE MURAMIC ACID RESIDUE.  
CC [1]- SUBUNIT: MONOMER.  
CC [1]- SUBCELLULAR LOCATION: ATTACHED TO THE OUTER MEMBRANE BY A LIPID  
CC ANCHOR AND EXPOSED TO THE PERIPLASMIC SIDE (PROBABLE).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; U18785; AAB60080.1; -  
DR EMBL; AE000354; AAC5743.1; -  
DR EMBL; J90892; CAB22492.1; -  
DR EMBL; J02708; -; NOT\_ANNOTATED\_CDS.  
DR PDB; 1LTM; 11-NOV-98  
DR ECOCENE; EGI2699; MTB.  
DR PROSITE; PS00013; PROKAR\_LIPOPROTEIN; 1.  
KW Cell wall; Hydrolase; Glycosidase; Signal; Lipoprotein;  
KW Outer membrane; Multigene family; 3D-structure.  
FT SIGNAL 1 18  
FT CHAIN 19 361  
FT MEMBRANE-BOUND LYTIC MUREIN  
FT TRANSGLYCOSYLASE B.  
FT N-ACYL DIGLYCERIDE (PROBABLE).  
SQ SEQUENCE 361 AA; 40256 MW; 93AF5C59 CRC32;  
Query Match 73.3%; Score 55; DB 1; Length 361;  
Best Local Similarity 55.6%; Pred. No. 2.62e+00;  
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;  
Db 327 YGLPNEYTTI 335  
QY 1 FAMPNEYTL 9  
RESULT 3  
ID NF1\_CHICK STANDARD; PRT; 270 AA.  
AC F35608;  
DT 01-JUN-1994 (Rel. 29, Created)  
DT 01-JUN-1994 (Rel. 29, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE NEUROFIBROMIN (NEUROFIBROMATOSIS-RELATED PROTEIN NF-1) (FRAGMENT).  
GN NF1.  
OS Gallus gallus (Chicken).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Archosauria; Aves;  
OC Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-BRAIN;  
RX MEDLINE; 93282908.  
RA SCHAPER G.L., CIMENT G., STOCKER K.M., BAIZER L.;  
RT "Analysis of the sequence and embryonic expression of chicken  
RT neurofibromin mRNA.";  
RL Mol. Chem. Neuropathol. 18:267-278(1993).  
CC [1]- FUNCTION: STIMULATES THE GTPASE ACTIVITY OF RAS. NF1 SHOWS GREATER  
CC AFFINITY FOR RAS GAP, BUT LOWER SPECIFIC ACTIVITY. THUS IT MAY BE  
CC A REGULATOR OF RAS ACTIVITY.  
CC [1]- SIMILARITY: TO OTHER RAS GTPASE-ACTIVATING PROTEINS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; S62087; AAB27069.1; -  
DR PROSITE; PS00509; RAS\_GTPASE\_ACTIV\_1; PARTIAL.  
DR PROSITE; PS00018; RAS\_GTPASE\_ACTIV\_2; PARTIAL.  
KW GTPase activation.  
FT NON\_TER 1 1  
FT NON\_TER 270 270  
SQ SEQUENCE 270 AA; 30753 MW; 653E2C8C CRC32;  
Query Match 72.0%; Score 54; DB 1; Length 270;  
Best Local Similarity 55.6%; Pred. No. 4.06e+00;





Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Fri Apr 14 23:31:32 2000; MasPar time 4.60 Seconds  
58.399 Million cell updates/sec  
Tabular output not generated.

Title: >US-08-452-843-6  
Description: (1-9) from US08452843.pep  
Perfect Score: 75  
Sequence: 1 FAMPNFYTL 9

Scoring table: PAM 150  
Gap 15

Searched: 82229 seqs, 29864866 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: swiss-prot38  
1:swissprot

Statistics: Mean 25.139; Variance 33.172; scale 0.758

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	56	74.7	195	1	YEH7 YEAST	1.68e+00
2	55	73.3	361	1	MLTB_ECOLI	2.62e+00
3	54	72.0	270	1	NFL1 CHICK	4.08e+00
4	54	72.0	433	1	POP2 YEAST	4.06e+00
5	54	72.0	753	1	PMT3 YEAST	4.06e+00
6	54	72.0	2839	1	NF1_HUMAN	4.06e+00
7	54	72.0	2841	1	NF1_MOUSE	4.06e+00
8	52	69.3	333	1	YH85 YEAST	9.61e+00
9	51	68.0	201	1	Y506_RICPR	1.47e+01
10	51	68.0	354	1	TYRA_LACLA	1.47e+01
11	51	68.0	869	1	AMPN_ECOLI	1.47e+01
12	50	66.7	169	1	NUGC MARPO	2.22e+01
13	50	66.7	198	1	PEMT RAT	2.22e+01
14	50	66.7	246	1	SPSI_BACSU	2.22e+01
15	50	66.7	381	1	Y926_HELPY	2.22e+01
16	50	66.7	410	1	YE28_CAEEL	2.22e+01
17	50	66.7	937	1	YNM3 YEAST	2.22e+01
18	49	65.3	700	1	RIRL_BACSU	3.34e+01
19	49	65.3	728	1	MYBA_XENLA	3.34e+01
20	49	65.3	844	1	HEXA_STRPN	3.34e+01
21	49	65.3	851	1	MUTS_RICPR	3.34e+01
22	48	64.0	375	1	ACT_GIALA	5.00e+01
23	48	64.0	402	1	PAIL_BOVIN	5.00e+01

24	48	64.0	463	1	UHPT_ECOLI	HEXOSE PHOSPHATE TRANS	5.00e+01
25	48	64.0	463	1	UHPT_SALTY	HEXOSE PHOSPHATE TRANS	5.00e+01
26	48	64.0	468	1	YOPH_YEREN	PROTEIN-TYROSINE PHOSP	5.00e+01
27	48	64.0	468	1	YOPH_YEREN	PROTEIN-TYROSINE PHOSP	5.00e+01
28	48	64.0	477	1	PEN3_ADECC	PENTON PROTEIN (VIRION	5.00e+01
29	48	64.0	532	1	PCKC_ANASU	PHOSPHOENOLPYRUVATE CA	5.00e+01
30	48	64.0	543	1	IEFS_HUMAN	TRANSFORMATION-SENSIT	5.00e+01
31	48	64.0	758	1	PMT2 YEAST	DOLICHYL-PHOSPHATE-MAN	5.00e+01
32	48	64.0	926	1	CHS2_SCHPO	CHITIN SYNTHASE 2 (EC	5.00e+01
33	48	64.0	3005	1	POIG_TVMV	GENOME POLYPROTEIN [CO	5.00e+01
34	47	62.7	122	1	YBEC_ECOLI	HYPOTHETICAL 14.2 KD P	7.43e+01
35	47	62.7	219	1	CAT_ECOLI	CHLORAMPHENICOL ACETYL	7.43e+01
36	47	62.7	246	1	Y181_METJA	HYPOTHETICAL PROTEIN M	7.43e+01
37	47	62.7	253	1	PG3_MASLA	PHYCOBILISOME ROD-CORE	7.43e+01
38	47	62.7	355	1	PUR5_ARATH	PHOSPHORIBOSYLFORMYLGL	7.43e+01
39	47	62.7	377	1	PVRC_ARATH	DIHYDROOROTASE PRECURS	7.43e+01
40	47	62.7	402	1	PAIL_RAT	PLASMINOGEN ACTIVATOR	7.43e+01
41	47	62.7	482	1	YPTL_CAEEL	HYPOTHETICAL 34.7 KD P	7.43e+01
42	47	62.7	597	1	SYK_AQUAE	LYSYL-TRNA SYNTHETASE	7.43e+01
43	47	62.7	685	1	MDL1_CANAL	ATP-DEPENDENT PERMEASE	7.43e+01
44	47	62.7	949	1	YMP9_YEAST	PUTATIVE 109.8 KD TRAN	7.43e+01
45	47	62.7	1324	1	MSH6_ARATH	DNA MISMATCH REPAIR PR	7.43e+01

ALIGNMENTS

RESULT 1	STANDARD;	PRT;	195 AA.
ID YEH7 YEAST			
AC P39978;			
DT 01-FEB-1995 (Rel. 31, Created)			
DT 01-FEB-1995 (Rel. 31, Last sequence update)			
DE 01-FEB-1995 (Rel. 31, Last annotation update)			
DE HYPOTHETICAL 21.7 KD PROTEIN IN HXT8-CAN1 INTERGENIC REGION.			
GN YEL067C.			
OS Saccharomyces cerevisiae (Baker's yeast).			
OC Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;			
OC Saccharomycetaceae; Saccharomycetes.			
RN [1]			
RP SEQUENCE FROM N.A.			
RC STRAIN-S288C / AB972;			
RA DIETRICH F.S., MULLIGAN J.T., HENNESSEY K.M., ALLEN E., ARAUJO R.,			
RA AVILES E., BERNO A., BRENNAN T., CARPENTER J., CHEN E., CHERRY J.M.,			
RA CHUNG E., DUNCAN M., GUZMAN E., HARTZELL G., HUNICKE-SMITH S.,			
RA HYMAN R., KAYSER A., KOMP C., LASHKARI D., LEW H., LIN D.,			
RA MOSEDALE D., NAKAHARA K., NAMATH A., NORGREN R., OEFNER P., OH C.,			
RA PETEL F.X., ROBERTS D., SEHL P., SCHRAMM S., SHOGREN T., SMITH V.,			
RA TAYLOR P., WEI Y., YELTON M., BORSTEIN D., DAVIS R.W.;			
RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.			
CC -----			
CC This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC the European Bioinformatics Institute. There are no restrictions on its			
CC use by non-profit institutions as long as its content is in no way			
CC modified and this statement is not removed. Usage by and for commercial			
CC entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>			
CC or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
CC -----			
DR EMBL: U18795; AAB65020.1; -			
KW Hypothetical protein.			
SQ SEQUENCE 195 AA; 21721 MW; 2EA97A20 CRC32;			
Query Match 74.7%; Score 56; DB 1; Length 195;			
Best Local Similarity 55.6%; Pred. No. 1.68e+00;			
Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;			
Db 17 FDMPTFFVL 25			
QY 1 FAMPNFYTL 9			
RESULT 2			
ID MLTB_ECOLI	STANDARD;	PRT;	361 AA.
AC P41052;			



THIS PAGE BLANK (uspro)

Mon Apr 17 08:20:56 2000

US-08-452-843-6.rpr

Page 7

OY 1 FAMPNYTL 9

Search completed: Fri Apr 14 23:31:15 2000  
Job time : 10 secs.

```
##molecule_type mRNA
##residues 1096-1569,'TPPEPET', ##label MA3
##cross-references GB:M61213; NID:g189162; PID:g189163
##note
this clone includes an epitope tag at the 3' end
encoding the sequence TPPEPET, not part of dystrophin
but recognized by the monoclonal antibody KT3

REFERENCE
#authors Nishi, T.; Lee, P.S.; Oka, K.; Levin, V.A.; Tanase, S.;
Morino, Y.; Saya, H.
#journal Oncogene (1991) 6:1555-1559
#title Differential expression of two types of the neurofibromatosis
type 1 (NF1) gene transcripts related to neuronal
differentiation.
#cross-references MUID:92019823
#accession I58356
#status translated from GB/EMBL/DBJ
#molecule_type mRNA
##residues 1168-1545 ##label RES
##cross-references GB:M60915; NID:g189159; PID:g189160
GENETICS
#gene GDB:NF1
#map_position 17q11.2-17q11.2
#introns 1370/3
#note the list of introns is incomplete
CLASSIFICATION #superfamily ras-specific GAP catalytic domain homology
KEYWORDS alternative splicing; tumor suppressor
FEATURE
1235-1449 #domain ras-specific GAP catalytic domain homology
#label GAP
SUMMARY
#length 2818 #molecular_weight 317030 #checksum 2858
Query Match 72.0%; Score 54; DB 2; Length 2818;
Best Local Similarity 55.6%; Pred. No. 1.04e+01;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 2135 FSLPKFYLL 2143
|:::|
QY 1 FAMPNEYTL 9

RESULT 13
ENTRY JCS196 #type complete
TITLE neurofibromin I - rat
ORGANISM #formal_name Rattus norvegicus #common_name Norway rat
DATE 20-Feb-1997 #sequence_revision 27-Feb-1997 #text_change
10-Sep-1997
ACCESSIONS JCS196
REFERENCE JCS196
#authors Suzuki, H.; Takahashi, K.; Yasumoto, K.; Fuse, N.; Shibahara,
S.
#journal J. Biochem. (1996) 120:1048-1054
#title Differential tissue-specific expression of neurofibromin
isoform mRNAs in rat.
#accession JCS196
#status preliminary; nucleic acid sequence not shown
##molecule_type mRNA
##residues 1-2820 ##label SUZ
##cross-references DBJ:D45201; NID:g1841313; PID:g1841314
COMMENT This protein contains a GTPase-activating protein-related domain
which is responsible for the stimulatory effect of neurofibromin
on the tyrosinase promoter activity.
CLASSIFICATION #superfamily ras-specific GAP catalytic domain homology
FEATURE
1177-1436 #domain GTPase-activating protein related #status
1237-1451 #domain ras-specific GAP catalytic domain homology
#label GAP
SUMMARY
#length 2820 #molecular_weight 317080 #checksum 6628
Query Match 72.0%; Score 54; DB 2; Length 2820;
Best Local Similarity 55.6%; Pred. No. 1.04e+01;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
```

```
Db 2137 FSLPKFYLL 2145
|:::|
QY 1 FAMPNEYTL 9

RESULT 14
ENTRY I54352 #type fragment
TITLE neurofibromin - mouse (fragment)
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE 02-Aug-1996 #sequence_revision 02-Aug-1996 #text_change
25-Apr-1997
ACCESSIONS I54352
REFERENCE I54352
#authors Bernards, A.; Snijders, A.J.; Hannigan, G.E.; Murthy, A.E.;
Gusella, J.F.
#journal Hum. Mol. Genet. (1993) 2:645-650
#title Mouse neurofibromatosis type 1 cDNA sequence reveals high
degree of conservation of both coding and non-coding mRNA
segments.
#cross-references MUID:93357730
#accession I54352
#status preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
##residues 1-2825 ##label RES
##cross-references GB:L10370; NID:g309452; PID:g309453
GENETICS
#gene NF1
CLASSIFICATION #superfamily ras-specific GAP catalytic domain homology
FEATURE
1221-1456 #domain ras-specific GAP catalytic domain homology
#label GAP
SUMMARY
#length 2825 #checksum 6076
Query Match 72.0%; Score 54; DB 2; Length 2825;
Best Local Similarity 55.6%; Pred. No. 1.04e+01;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 2142 FSLPKFYLL 2150
|:::|
QY 1 FAMPNEYTL 9

RESULT 15
ENTRY JT0366 #type fragment
TITLE hypothetical protein 1 - bullfrog mitochondrion (SGC1)
(fragment)
ORGANISM #formal_name mitochondrion Rana catesbeiana #common_name
bullfrog
DATE 23-Oct-1992 #sequence_revision 23-Oct-1992 #text_change
31-Dec-1993
ACCESSIONS JT0366
REFERENCE JT0366
#authors Fujii, H.
#journal Nichidaishi (1987) 54:59-71
#title Cloning of the entire mitochondrial genome of Rana
catesbeiana and nucleotide sequencing of the URF2 and its
flanking genes.
#accession JT0366
##molecule_type DNA
##residues 1-61 ##label FUJ
GENETICS
#genome mitochondrion
#genetic_code SGC1
KEYWORDS mitochondrion
SUMMARY
#length 61 #checksum 9775
Query Match 70.7%; Score 53; DB 2; Length 61;
Best Local Similarity 77.8%; Pred. No. 1.55e+01;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 25 FAMINLYTL 33
|:::|
```

```

#cross-references MUID:96158055
#accession S60414
##status nucleic acid sequence not shown
##molecule_type DNA
##residues 1-396,'H',398-566,'N',568-753 ##label IMX
##cross-references EMBL:X83797; NID:g633651; PID:g633652
REFERENCE
#authors Pearson, B.M.; Hernando, Y.; Kalogeropoulos, A.; Schweizer,
M.
#submission submitted to the Protein Sequence Database, July 1996
#accession S67227
##molecule_type DNA
##residues 1-753 ##label PEW
##cross-references EMBL:275229; NID:g1420703; PID:e252150; PID:g1420704;
MIPS:YOR321W
##experimental_source strain S288C
REFERENCE
#authors Pearson, B.M.; Hernando, Y.; Payne, J.; Wolf, S.S.;
Kalogeropoulos, A.; Schweizer, M.
#journal Yeast (1996) 12:1021-1031
#title Sequencing of a 35.71 kb DNA segment on the right arm of
yeast chromosome XV reveals regions of similarity to
chromosomes I and XIII.
#accession S72001
##status nucleic acid sequence not shown; translation not shown
##molecule_type DNA
##residues 1-753 ##label PEF
##cross-references EMBL:X90565; NID:g940836; PID:g940852
#note the nucleotide sequence was submitted to the EMBL Data
Library, August 1995
GENETICS
#gene SGD:PMT3
##cross-references SGD:S0005848; MIPS:YOR321W
#map_position 15R
CLASSIFICATION #superfamily dolichyl-phosphate-mannose--protein
mannosyltransferase
endoplasmic reticulum; glycosyltransferase;
hexosyltransferase; transmembrane protein
FEATURE
55-71 #domain transmembrane #status predicted #label TM1\
166-182 #domain transmembrane #status predicted #label TM2\
192-208 #domain transmembrane #status predicted #label TM3\
239-255 #domain transmembrane #status predicted #label TM4\
284-300 #domain transmembrane #status predicted #label TM5\
607-623 #domain transmembrane #status predicted #label TM6\
640-656 #domain transmembrane #status predicted #label TM7\
704-720 #domain transmembrane #status predicted #label TM8\
SUMMARY
#length 753 #molecular_weight 86322 #checksum 3415
Query Match 72.0%; Score 54; DB 2; Length 753;
Best Local Similarity 55.6%; Pred. No. 1.04e+01;
Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
Db 641 FVMAAFYPL 649
| | | | |
| | | | |
Qy 1 FAMPNFYTL 9
RESULT 12
ENTRY
#type complete
TITLE neurofibromatosis-related protein NF1 - human
ALTERNATE_NAMES GTPase activating protein homolog NF1; neurofibromin
ORGANISM #formal_name Homo sapiens #common_name man
DATE 10-Feb-1995 #sequence_revision 10-Feb-1995 #text_change
20-Mar-1998
ACCESSIONS B55282; A55282; A35879; A35605; A35910; A35222; A36297;
158356
REFERENCE
#authors Marchuk, D.A.; Saulino, A.M.; Tavakoli, R.; Swaroop, M.;
Wallace, M.R.; Andersen, L.B.; Mitchell, A.L.; Gutmann,
D.H.; Boguski, M.; Collins, F.S.
#journal Genomics (1991) 11:931-940
#title cDNA cloning of the type 1 neurofibromatosis gene: complete
sequence of the NF1 gene product.
#cross-references MUID:92147138
#accession B55282
##status not compared with conceptual translation
##molecule_type mRNA
##residues 1-2818 ##label MAR
##cross-references GB:M82814; NID:g189164; PID:g189165
#note sequence extracted from NCBI backbone (NCBI:80176)
#accession A55282
##status preliminary
##molecule_type mRNA
##residues 1-334 ##label MA2
#note sequence extracted from NCBI backbone (NCBIN:80169,
NCBI:P:80172)
REFERENCE
#authors Xu, G.; O'Connell, P.; Viskochil, D.; Cawthon, R.; Robertson,
M.; Culver, M.; Dunn, D.; Stevens, J.; Gesteland, R.;
White, R.; Weiss, R.
#journal Cell (1990) 62:599-608
#title The neurofibromatosis type 1 gene encodes a protein related
to GAP.
#cross-references MUID:90335969
#accession A35879
##status preliminary
##molecule_type mRNA
##residues 335-495,'I',497-1555,'H',1556-2818 ##label XUA
##cross-references GB:M38106; GB:M57449; NID:g189169; PID:g189170
A35605
#authors Cawthon, R.M.; Weiss, R.; Xu, G.; Viskochil, D.; Culver, M.;
Stevens, J.; Robertson, M.; Dunn, D.; Gesteland, R.;
O'Connell, P.; White, R.
#journal Cell (1990) 62:193-201
#title A major segment of the neurofibromatosis type 1 gene: cDNA
sequence, genomic structure, and point mutations.
#cross-references MUID:90304909
#accession A35605
##status preliminary
##molecule_type mRNA
##residues 1585-2687 ##label CAW
##cross-references EMBL:M38107; EMBL:M57449
A35910
#authors Cawthon, R.M.; Weiss, R.; Xu, G.; Viskochil, D.; Culver, M.;
Stevens, J.; Robertson, M.; Dunn, D.; Gesteland, R.;
O'Connell, P.; White, R.
#journal Cell (1990) 62:608b
#accession A35910
##status preliminary; nucleic acid sequence not shown; not
compared with conceptual translation
##molecule_type mRNA
##residues 2688-2818 ##label CA2
#authors Wallace, M.R.; Marchuk, D.A.; Andersen, L.B.; Letcher, R.;
Odeh, H.M.; Saulino, A.M.; Fountain, J.W.; Brereton, A.;
Nicholson, J.; Mitchell, A.L.; Brownstein, B.H.; Collins,
F.S.
#journal Science (1990) 249:181-186
#title Type 1 neurofibromatosis gene: identification of a large
transcript disrupted in three NF1 patients.
#cross-references MUID:90319792
#accession A35222
##status preliminary
##molecule_type mRNA
##residues 2209-2818 ##label WAL
##cross-references GB:M60496; NID:g189157; PID:g189158; GB:M49193
A36297
#authors Martin, G.A.; Viskochil, D.; Bollag, G.; McCabe, P.C.;
Crosier, W.J.; Haubruck, H.; Conroy, L.; Clark, R.;
O'Connell, P.; Cawthon, R.M.; Innis, M.A.; McCormick, F.
#journal Cell (1990) 63:843-849
#title The GAP-related domain of the neurofibromatosis type 1 gene
product interacts with ras p21.
#cross-references MUID:91029515
#accession A36297

```

```
C.G.
#journal Nature (1998) 396:133-140
#title The genome sequence of Rickettsia prowazekii and the origin
#accession A71733
##status preliminary; nucleic acid sequence not shown;
      translation not shown
##molecule_type DNA
##residues 1-339 ##label AND
##cross-references GB:AJ235270; GB:AJ235269; NID:g3860572; PID:el342523;
##experimental_source strain Madrid E
GENETICS
#gene cydB; RP217
#length 339 #molecular-weight 38029 #checksum 490
SUMMARY
Query Match 72.0%; Score 54; DB 2; Length 339;
Best Local Similarity 66.7%; Pred. No. 1.04e+01;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Db 230 FSMPIIYL 238
Qy 1 FAMPNFTYL 9
RESULT 9
ENTRY #type complete
TITLE POP2 protein - yeast (Saccharomyces cerevisiae)
ALTERNATE_NAMES protein N3470; protein YNR052c
ORGANISM #formal_name Saccharomyces cerevisiae
DATE 27-Apr-1996 #sequence_revision 03-May-1996 #text_change
ACCESSIONS S63383; S35997; S35996; S36929; S27438
REFERENCE S63346
#authors Pohl, T.M.
#submission submitted to the Protein Sequence Database, April 1996
#accession S63383
##molecule_type DNA
##residues 1-433 ##label POH
##cross-references EMBL:Z71667; NID:gl302567; PID:e239839; PID:gl302568;
      ##experimental_source MIPS:YNR052c
S35996
#authors Sakai, A.; Chibazakura, T.; Shimizu, Y.; Hishinuma, F.
#journal Nucleic Acids Res. (1992) 20:6227-6233
#title Molecular analysis of POP2 gene, a gene required for
      glucose-derepression of gene expression in Saccharomyces
      cerevisiae
#cross-references MIMD:93117094
#accession S35997
##status nucleic acid sequence not shown
##residues 1-80,82-411,'M',413-433 ##label SAK
##cross-references GB:DL2807
##experimental_source strain S288C
#accession S35996
##status nucleic acid sequence not shown
##residues 1-40,'Q',42-91,'QQQQQQQQQQQQQQQ',92-111,117-277,'S',
      279-433 ##label SAW
##cross-references GB:DL2808
##experimental_source strain A364A
REFERENCE S36929
#authors Sakai, A.; Chibazakura, T.; Shimizu, Y.; Hishinuma, F.
#submission submitted to the EMBL Data Library, August 1992
#accession S36929
##molecule_type DNA
##residues 1-91,'QQQQQQQQQQQQQQQQQ',92-111,117-277,'S',279-433
      ##label SA2
##cross-references GB:DL2808; NID:g218463; PID:d1002742; PID:g218463
##experimental_source strain A364A
REFERENCE S27437
#authors Cusick, M.E.
```

```
#submission submitted to the EMBL Data Library, March 1992
#accession S27438
##molecule_type DNA
##residues 213-433 ##label CUS
##cross-references EMBL:M88607; NID:gl72079; PID:gl72080
GENETICS
#gene SGP:POP2; CAF1
#map_position 14R
#cross-references SGD:S0005335; MIPS:YNR052c
SUMMARY
#length 433 #molecular-weight 49682 #checksum 2617
Query Match 72.0%; Score 54; DB 2; Length 433;
Best Local Similarity 85.7%; Pred. No. 1.04e+01;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Db 338 MPNFYDL 344
Qy 3 MPNFYTL 9
RESULT 10
ENTRY #type fragment
TITLE neurofibromatosis-related protein NF1 - mouse (fragment)
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE 18-Feb-1994 #sequence_revision 10-Nov-1995 #text_change
ACCESSIONS S11510
REFERENCE S11510
#authors Buchberg, A.M.; Cleveland, L.S.; Jenkins, N.A.; Copeland,
      N.G.
#journal Nature (1990) 347:291-294
#title Sequence homology shared by neurofibromatosis type-1 gene and
      IRA-1 and IRA-2 negative regulators of the RAS cyclic AMP
      pathway.
#cross-references MIMD:90384569
#accession S11510
##status preliminary
##molecule_type mRNA
##residues 1-621 ##label BUC
SUMMARY #length 621 #checksum 587
Query Match 72.0%; Score 54; DB 2; Length 621;
Best Local Similarity 55.6%; Pred. No. 1.04e+01;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
Db 209 FSLPKFYLL 217
Qy 1 FAMPNFTYL 9
RESULT 11
ENTRY #type complete
TITLE dolichyl-phosphate-mannose--protein mannosyltransferase (EC
      2.4.1.109) PMT3 - yeast (Saccharomyces cerevisiae)
ALTERNATE_NAMES protein O6148; protein YOR321w
ORGANISM #formal_name Saccharomyces cerevisiae
DATE 13-Jan-1996 #sequence_revision 01-Mar-1996 #text_change
ACCESSIONS S58331; S60414; S67227; S72001; S51283
REFERENCE S58318
#authors Pearson, B.M.; Hernando, Y.; Wolf, S.S.; Kalogeropoulos, A.;
      Schweizer, M.
#submission submitted to the EMBL Data Library, August 1995
#accession S58331
##molecule_type DNA
##residues 1-753 ##label PEA
##cross-references EMBL:X90565; NID:g940836; PID:g940852
      S60414
#authors Immervoll, T.; Gentzsch, M.; Tanner, W.
#journal Yeast (1995) 11:1345-1351
#title PMT3 and PMT4, two new members of the
      protein-O-mannosyltransferase gene family of Saccharomyces
      cerevisiae.
```

C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karamata, D.; Kashara, Y.; Klaerr-Blanchard, M.; Klein, C.; Kobayashi, Y.; Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois, S.; Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno, M.; Moestl, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly, M.; Ogawa, K.; Ogilwa, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle, D.; Porvolik, S.; Prescott, G.; A.M.; Presseau, E.; Pujic, P.; Purnelle, B.; Rapoport, G.; Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon, E.; Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Setor, S.J.; Serror, P.; Shin, B.S.; Soldo, B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Takahashi, H.; Takemaru, K.; Takeuchi, M.; Tanakoshi, A.; Tanaka, T.; Terpestra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.; Vandenbol, M.; Vannier, F.; Vassarotti, A.; Viari, A.; Wambutt, R.; Wedler, E.; Wedler, H.; Weitzenegger, T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.; Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.

#journal Nature (1997) 390:249-256  
#title The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.  
#cross-references MUID:98044033  
#accession A69927  
#status preliminary; nucleic acid sequence not shown;  
translation not shown

##molecule\_type DNA  
##residues 1-786 ##label KUN  
##cross-references GB:Z99114; GB:AL009136; NID:g2634230; PID:e1185477;  
##experimental\_source strain 168

GENETICS  
#gene yosO  
#summary #length 786 #molecular-weight 90101 #checksum 4364

Query Match 73.38; Score 55; DB 2; Length 786;  
Best Local Similarity 66.78; Pred. No. 6.99e+00;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 578 FMVNFYSL 586  
| : | : | : |  
QY 1 FAMPNFYTL 9

RESULT 6  
ENTRY #type complete  
TITLE hypothetical protein MTH115 - Methanobacterium  
ORGANISM thermoautotrophicum (strain Delta H)  
#formal\_name Methanobacterium thermoautotrophicum  
DATE 05-Dec-1997 #sequence\_revision 03-Dec-1997 #text\_change  
ACCESSIONS B69020  
REFERENCE B69020  
#authors Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.; Bashirzadeh, R.; Blakely, D.; Cook, R.; Gilbert, K.; Harrison, D.; Hoang, R.; Keagle, P.; Lumm, W.; Pothier, B.; Qiu, D.; Spadafora, R.; Vicaire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiawani, N.; Caruso, A.; Bush, D.; Safer, H.; Patwell, D.; Prabakar, S.; McDougall, S.; Shimer, G.; Goyal, A.; Pietrokovski, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.  
#journal J. Bacteriol. (1997) 179:7135-7155  
#title Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: functional analysis and comparative genomics.  
#cross-references MUID:98037514  
#accession B69020  
#status preliminary; nucleic acid sequence not shown;

##molecule\_type DNA  
##residues 1-189 ##label MTH  
##cross-references GB:AE000801; GB:AE000666; NID:g2621145; PID:g2621154  
##experimental\_source strain Delta H

GENETICS  
#gene MTH115  
#start\_codon TTG  
#summary #length 189 #molecular-weight 21688 #checksum 2732

Query Match 72.08; Score 54; DB 2; Length 189;  
Best Local Similarity 66.78; Pred. No. 1.04e+01;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 100 FTMPSYTL 108  
| : | : | : |  
QY 1 FAMPNFYTL 9

RESULT 7  
ENTRY #type complete  
TITLE hypothetical protein PH1547 - Pyrococcus horikoshii  
ORGANISM Pyrococcus horikoshii  
#formal\_name Pyrococcus horikoshii  
DATE 14-Aug-1998 #sequence\_revision 14-Aug-1998 #text\_change  
ACCESSIONS C71032  
REFERENCE C71032  
#authors Kawabayashi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Sekine, M.; Baba, S.; Kosugi, H.; Hosoyama, A.; Negai, Y.; Sakai, M.; Ogura, K.; Otsuka, R.; Nakazawa, H.; Takamiya, M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kishida, N.; Oguchi, A.; Aoki, K.; Yoshizawa, T.; Nakamura, Y.; Robb, F.T.; Horikoshi, K.; Masuchi, Y.; Shizuya, H.; Kikuchi, H.

DNA Res. (1998) 5:55-76  
Complete sequence and gene organization of the genome of a hyper-thermophilic archaeobacterium, *Pyrococcus horikoshii* OT3.  
#cross-references MUID:98344137  
#accession C71032  
#status preliminary; nucleic acid sequence not shown;  
translation not shown

##molecule\_type DNA  
##residues 1-206 ##label KAW  
##cross-references GB:AP000006; NID:g3236133; PID:d1031602; PID:g3257976  
##experimental\_source strain Ot3  
##note this accession replaces an interim accession for a sequence replaced by GenBank

GENETICS  
#gene PH1547  
#summary #length 206 #molecular-weight 24035 #checksum 1016

Query Match 72.08; Score 54; DB 2; Length 206;  
Best Local Similarity 44.48; Pred. No. 1.04e+01;  
Matches 4; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 14 FSMATYSL 22  
| : | : | : |  
QY 1 FAMPNFYTL 9

RESULT 8  
ENTRY #type complete  
TITLE cytochrome D ubiquinol oxidase chain II (cydB) RP217 -  
ORGANISM Rickettsia prowazekii  
#formal\_name Rickettsia prowazekii  
DATE 21-Nov-1998 #sequence\_revision 21-Nov-1998 #text\_change  
ACCESSIONS A71733  
REFERENCE A71630  
#authors Andersson, S.G.E.; Zomorodipour, A.; Andersson, J.O.; Sichteritz-Ponten, T.; Alsmark, U.C.M.; Podowski, R.M.; Naeslund, A.K.; Eriksson, A.S.; Winkler, H.H.; Kurland,

Query Match 72.08; Score 54; DB 2; Length 206;  
Best Local Similarity 44.48; Pred. No. 1.04e+01;  
Matches 4; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

```

##residues      1-195 ##label DIE
##cross-references EMBL:U18795; NID:g603241; PID:g603251; MIPS:YEL067c
GENETICS
#map_position 5L
SUMMARY
#length 195 #molecular-weight 21721 #checksum 6000

Query Match      74.7%; Score 56; DB 2; Length 195;
Best Local Similarity 53.6%; Pred. No. 4.67e+00;
Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Db 17 FDMPTFFVL 25
| | | | |
QY 1 FAMPNFYTL 9

RESULT 3
ENTRY C69293 #type complete
TITLE C4-dicarboxylate transporter (mael) homolog - Archaeoglobus
ORGANISM fulgidus
DATE #formal_name Archaeoglobus fulgidus
05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change
ACCESSIONS C69293
REFERENCE A69250
#authors Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson,
K.E.; Ketchum, K.A.; Dodson, R.J.; Gwinn, M.; Hickey, E.K.;
Peterson, J.D.; Richardson, D.L.; Kierlavage, A.R.; Graham,
D.E.; Kyrpides, N.C.; Fleischmann, R.D.; Quackenbush, J.;
Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.;
Dougherty, B.A.; McKenny, K.; Adams, M.D.; Loftus, B.;
Peterson, S.; Reich, C.I.; McNeil, L.K.; Badger, J.H.;
Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman,
J.F.; McDonald, L.; Utterback, T.; Cotton, M.D.; Spriggs,
T.; Artach, P.; Kaine, B.P.; Sykes, S.M.; Sadow, P.W.;
D'Andrea, K.P.; Bowman, C.; Fujii, C.; Garland, S.A.;
Mason, T.M.; Olsen, G.J.; Fraser, C.M.; Smith, H.O.; Woese,
C.R.; Venter, J.C.
#journal Nature (1997) 390:364-370
#title The complete genome sequence of the hyperthermophilic,
sulfate-reducing archaeon Archaeoglobus fulgidus.
#cross-references MUID:98049343
#accession C69293
#status preliminary; nucleic acid sequence not shown;
translation not shown
#molecule_type DNA
##residues 1-319 ##label KLE
##cross-references GB:AE001080; GB:AE000782; NID:g2689403; PID:g2650284;
TIGR:AF0347
SUMMARY
#length 319 #molecular-weight 35267 #checksum 7060

Query Match      74.7%; Score 56; DB 2; Length 319;
Best Local Similarity 66.7%; Pred. No. 4.67e+00;
Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Db 68 FVMGNFYPL 76
| | | | |
QY 1 FAMPNFYTL 9

RESULT 4
ENTRY A65050 #type complete
TITLE membrane-bound lytic transglycosylase (EC 3.2.1.-) B
ALTERNATE_NAMES precursor - Escherichia coli
ORGANISM mltB protein
DATE #formal_name Escherichia coli
12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change
ACCESSIONS A65050; S65868; S77642
REFERENCE A64720
#authors Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.;
Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.;
Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.;
Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao,

```

```

Y.
#journal Science (1997) 277:1453-1462
#title The complete genome sequence of Escherichia coli K-12.
#cross-references MUID:97426617
#accession A65050
#status nucleic acid sequence not shown; translation not shown
#molecule_type DNA
##residues 1-361 ##label BLAT
##cross-references GB:AE000354; GB:U00096; NID:g2367149; PID:gl789053;
UWGP:b2701
#experimental_source strain K-12, substrain MG1655
REFERENCE S65868
#authors Dijkstra, A.J.; Hermann, F.; Keck, W.
#journal FEBS Lett. (1995) 366:115-118
#title Cloning and controlled overexpression of the gene encoding
the 35 kDa soluble lytic transglycosylase from Escherichia
coli.
#cross-references MUID:95309413
#accession S65868
#status preliminary; not compared with conceptual translation
#molecule_type DNA
##residues 1-34, 'A', 36-361 ##label DIJ
REFERENCE S77642
#authors Ehler, K.; Hoeltje, J.V.; Templin, M.F.
#journal Mol. Microbiol. (1995) 16:761-768
#title Cloning and expression of a murein hydrolase lipoprotein from
Escherichia coli.
#accession S77642
#status preliminary; nucleic acid sequence not shown;
translation not shown
#molecule_type DNA
##residues 1-361 ##label EHL
##cross-references EMBL:U18785; NID:g642537; PID:g642538
#note the nucleotide sequence was submitted to the EMBL Data
Library, December 1994
GENETICS
#gene mltB
#keywords glycosidase; hydrolase
FEATURE
1-18
19-361
SUMMARY
#length 361 #molecular-weight 40256 #checksum 6127

Query Match      73.3%; Score 55; DB 2; Length 361;
Best Local Similarity 55.6%; Pred. No. 6.99e+00;
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 327 YGLPNEYTI 335
QY 1 FAMPNFYTL 9

RESULT 5
ENTRY A69927 #type complete
TITLE ribonucleoside-diphosphate reductase (alph) homolog yosO -
Bacillus subtilis
ORGANISM #formal_name Bacillus subtilis
DATE 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change
24-Sep-1998
ACCESSIONS A69927
REFERENCE A89580
#authors Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.;
Alloni, G.; Azevedo, V.; Berto, M.G.; Bessieres, P.;
Bolotin, A.; Borchert, S.; Boriss, R.; Boursier, L.; Brans,
A.; Braun, M.; Brignell, S.C.; Bron, S.; Brouillet, S.;
Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
Choi, S.K.; Codani, J.J.; Connerton, I.F.; Cummings, N.J.;
Daniel, R.A.; Denizot, F.; Devine, K.M.; Duesterhoft, A.;
Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.;
Fabret, C.; Ferrari, E.; Foulger, D.; Fritz, C.; Fujita,
M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galleron, N.; Ghim,
S.Y.; Glaser, P.; Goffeau, A.; Golightly, E.J.; Grandi, G.;
Guiseppi, G.; Guy, B.J.; Hage, K.; Harech, J.; Harwood,

```





THIS PAGE BLANK (USPTO)

```

Db 125 YAMPEFQ 131
QY 1 FAMPNFQ 7

RESULT 15
ID TORA_SHEMA STANDARD; PRT; 829 AA.
AC 087948;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE TRIMETHYLAMINE-N-OXIDE REDUCTASE PRECURSOR (EC 1.6.6.9) (TMAO
  REDUCTASE) (TRIMETHYLAMINE OXIDASE).
GN TORA.
OS Shewanella massilia.
OC Bacteria; Proteobacteria; gamma subdivision; Alteromonadaceae;
OC Shewanella.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 99033056.
RA DOS SANTOS J.P., IOBBI-NIVOL C., COUILLAUD C., GIORDANO G.,
RA MEJEAN V.;
RT "Molecular analysis of the trimethylamine N-oxide (TMAO) reductase
RT respiratory system from a Shewanella species.";
RL J. Mol. Biol. 284:421-433(1998).
RN [2]
RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).
RX MEDLINE; 99033057.
RA CZJZEK M., DOS SANTOS J.P., POMMIER J., GIORDANO G., MEJEAN V.,
RA HASER R.;
RT "Crystal structure of oxidized trimethylamine N-oxide reductase from
RT Shewanella massilia at 2.5-A resolution.";
RL J. Mol. Biol. 284:435-447(1998).
CC -|- FUNCTION: REDUCES TRIMETHYLAMINE-N-OXIDE (TMAO) INTO
CC TRIMETHYLAMINE; AN ANAEROBIC REACTION COUPLED TO ENERGY-YIELDING
CC REACTIONS.
CC -|- CATALYTIC ACTIVITY: NADH + TRIMETHYLAMINE-N-OXIDE = NAD(+) +
CC TRIMETHYLAMINE + H(2)O.
CC -|- SUBCELLULAR LOCATION: PERIPLASMIC.
CC -|- SIMILARITY: BELONGS TO THE PROKARYOTIC MOLYBDOPTERIN-CONTAINING
CC OXIDOREDUCTASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AJ006085; CAA06851.1; -
DR PDB; 1TMO; 30-MAR-99.
DR PROSITE; PS00551; MOLYBDOPTERIN_PROK_1; FALSE_NEG.
DR PROSITE; PS00490; MOLYBDOPTERIN_PROK_2; 1.
DR PROSITE; PS00932; MOLYBDOPTERIN_PROK_3; 1.
DR PFAM; PF00384; molybdopterin; 1.
DR PFAM; PF01568; Molydop_binding; 1.
KW Oxidoreductase; NAD; Molybdenum; Periplasmic; Signal; 3D-structure.
FT SIGNAL 1 31
FT CHAIN 32 829 TRIMETHYLAMINE-N-OXIDE REDUCTASE.
SQ SEQUENCE 829 AA; 92362 MW; A5307DA9 CRC32;

Query Match 70.0%; Score 49; DB 1; Length 829;
Best Local Similarity 55.6%; Pred. No. 6.50e+00;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 593 FEMPDFATF 601
QY 1 FAMPNFQIL 9

```

```

RP SEQUENCE FROM N.A.
RA ANWARUL H.K., MORIYA S., BAUMANN P., YOSHIKAWA H., OGASAWARA N.;
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE: 98184963.
RA CLARK M.A., BAUMANN L., BAUMANN P.;
RT "Sequence analysis of a 34.7-kb DNA segment from the genome of
RT Buchnera aphidicola (endosymbiont of aphids) containing groEL, dnaA,
RT the atp operon, gldA, and rho.";
RL Curr. Microbiol. 36:158-163(1998).
CC -!- FUNCTION: INVOLVED IN THIOPHENE OXIDATION (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE ERA/THDF FAMILY OF GTP-BINDING
CC PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D85628; BAA12845.1; -
DR EMBL: AF008210; AAC38101.1; -
KW GTP-binding.
FT NP_BIND 225 232 GTP (BY SIMILARITY).
FT NP_BIND 272 276 GTP (BY SIMILARITY).
FT NP_BIND 338 341 GTP (BY SIMILARITY).
SQ SEQUENCE 456 AA; 51695 MW; 6A3ECD06 CRC32;

Query Match 70.0%; Score 49; DB 1; Length 456;
Best Local Similarity 55.6%; Pred. No. 6.50e+00;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 190 FINSFEKL 198
QY 1 FAMPNFQTL 9

RESULT 14
ID AMD3 HUMAN STANDARD; PRT; 767 AA.
AC Q01432;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE AMP DEAMINASE 3 (EC 3.5.4.6) (AMP DEAMINASE ISOFORM E).
GN AMPD3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 93042002.
RA YAMADA Y., GOTO H., OGASAWARA N.;
RT "Cloning and nucleotide sequence of the cDNA encoding human
RT erythrocyte-specific AMP deaminase.";
RL Biochim. Biophys. Acta 1171:125-128(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE: 93015995.
RA MAHNE-ZIZELMAN D.K., SABINA R.L.;
RT "Cloning of human AMP deaminase isoform E cDNAs. Evidence for a third
RT AMPD gene exhibiting alternatively spliced 5'-exons.";
RL J. Biol. Chem. 267:20866-20877(1992).
RN [3]
RP SEQUENCE FROM N.A.
RA MAHNE-ZIZELMAN D.K., EDDY R., SHOWS T.B., SABINA R.L.;
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA YAMADA Y., GOTO H., MURASE T., OGASAWARA N.;

```

```

RL Submitted (NOV-1994) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: AMP DEAMINASE PLAYS A CRITICAL ROLE IN ENERGY
CC METABOLISM.
CC -!- CATALYTIC ACTIVITY: AMP + H(2)O -> IMP + NH(3).
CC -!- PATHWAY: PURINE NUCLEOTIDE CYCLE.
CC -!- SUBUNIT: HOMOTETRAMER.
CC -!- ALTERNATIVE PRODUCTS: THREE FORMS OF AMP DEAMINASE 3 (1A, 1B AND
CC 1C) ARE PRODUCED BY ALTERNATIVE SPLICING OF THE SAME GENE.
CC -!- TISSUE SPECIFICITY: THREE ISOFORMS ARE PRESENT IN MAMMALS: AMP
CC DEAMINASE 1 IS THE PREDOMINANT FORM IN SKELETAL MUSCLE; AMP
CC DEAMINASE 2 PREDOMINATES IN SMOOTH MUSCLE, NON-MUSCLE TISSUE,
CC ERYTHRONIC MUSCLE AND UNDIFFERENTIATED MYOBLASTS; AMP DEAMINASE 3
CC IS FOUND IN ERYTHROCYTES.
CC -!- SIMILARITY: BELONGS TO THE ADENOSINE AND AMP DEAMINASES FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D12775; BAA02240.1; -
DR EMBL: M84720; AAA58365.1; -
DR EMBL: M84721; AAA58366.1; -
DR EMBL: M84722; AAA58367.1; -
DR EMBL: U29926; AAB60410.1; -
DR EMBL: U29929; AAB60410.1; JOINED.
DR EMBL: U29907; AAB60410.1; JOINED.
DR EMBL: U29909; AAB60410.1; JOINED.
DR EMBL: U29910; AAB60410.1; JOINED.
DR EMBL: U29911; AAB60410.1; JOINED.
DR EMBL: U29916; AAB60410.1; JOINED.
DR EMBL: U29917; AAB60410.1; JOINED.
DR EMBL: U29918; AAB60410.1; JOINED.
DR EMBL: U29922; AAB60410.1; JOINED.
DR EMBL: U29924; AAB60410.1; JOINED.
DR EMBL: U29925; AAB60410.1; JOINED.
DR EMBL: D31646; BAA06505.1; -
DR EMBL: D31633; BAA06505.1; JOINED.
DR EMBL: D31634; BAA06505.1; JOINED.
DR EMBL: D31635; BAA06505.1; JOINED.
DR EMBL: D31637; BAA06505.1; JOINED.
DR EMBL: D31638; BAA06505.1; JOINED.
DR EMBL: D31639; BAA06505.1; JOINED.
DR EMBL: D31640; BAA06505.1; JOINED.
DR EMBL: D31641; BAA06505.1; JOINED.
DR EMBL: D31642; BAA06505.1; JOINED.
DR EMBL: D31643; BAA06505.1; JOINED.
DR EMBL: D31644; BAA06505.1; JOINED.
DR EMBL: D31645; BAA06505.1; JOINED.
DR PIR: B45071; B45071.
DR PIR: S28149; S28149.
DR MIM: 102772; -.
DR PROSITE: PS00485; A_DEAMINASE; 1.
DR PFAM: PF00962; A_deaminase; 1.
KW Hydrolase; Nucleotide metabolism; Multigene family;
KW Alternative splicing.
FT ACT_SITE 377 377 POTENTIAL.
FT ACT_SITE 587 587 POTENTIAL.
FT ACT_SITE 663 663 POTENTIAL.
FT ACT_SITE 664 664 POTENTIAL.
FT VARSPPLIC 1 1 M -> MALSSPAEM (IN ISOFORM 1A).
FT VARSPPLIC 1 1 M -> MEPSAEM (IN ISOFORM 1C).
FT VARSPPLIC 208 767 MISSING (IN ISOFORM 1A).
FT VARSPPLIC 652 767 MISSING (IN ISOFORM 1C).
SQ SEQUENCE 767 AA; 88812 MW; 9D95DAB1 CRC32;

Query Match 70.0%; Score 49; DB 1; Length 767;
Best Local Similarity 71.4%; Pred. No. 6.50e+00;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

```

FT DOMAIN 860 874 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 875 897 POTENTIAL.  
FT DOMAIN 898 931 CYTOPLASMIC (POTENTIAL).  
FT DOMAIN 932 80 EGF-LIKE 1.  
FT DOMAIN 81 132 EGF-LIKE 2.  
FT DOMAIN 133 172 EGF-LIKE 3, CALCIUM-BINDING (POTENTIAL).  
FT DOMAIN 173 221 EGF-LIKE 4, CALCIUM-BINDING (POTENTIAL).  
FT DOMAIN 222 271 EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).  
FT DOMAIN 272 318 EGF-LIKE 6, CALCIUM-BINDING (POTENTIAL).  
FT DOMAIN 319 367 EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).  
FT SITE 506 508 CELL ATTACHMENT SITE (POTENTIAL).  
FT DISULFID 36 48 BY SIMILARITY.  
FT DISULFID 42 57 BY SIMILARITY.  
FT DISULFID 59 79 BY SIMILARITY.  
FT DISULFID 85 98 BY SIMILARITY.  
FT DISULFID 92 107 BY SIMILARITY.  
FT DISULFID 109 131 BY SIMILARITY.  
FT DISULFID 137 149 BY SIMILARITY.  
FT DISULFID 143 158 BY SIMILARITY.  
FT DISULFID 160 171 BY SIMILARITY.  
FT DISULFID 177 189 BY SIMILARITY.  
FT DISULFID 183 198 BY SIMILARITY.  
FT DISULFID 200 220 BY SIMILARITY.  
FT DISULFID 226 239 BY SIMILARITY.  
FT DISULFID 233 248 BY SIMILARITY.  
FT DISULFID 250 270 BY SIMILARITY.  
FT DISULFID 276 286 BY SIMILARITY.  
FT DISULFID 280 295 BY SIMILARITY.  
FT DISULFID 297 317 BY SIMILARITY.  
FT DISULFID 323 336 BY SIMILARITY.  
FT DISULFID 330 345 BY SIMILARITY.  
FT DISULFID 347 366 BY SIMILARITY.  
FT CARBOHYD 148 148 POTENTIAL.  
FT CARBOHYD 167 167 POTENTIAL.  
FT CARBOHYD 229 229 POTENTIAL.  
FT CARBOHYD 269 269 POTENTIAL.  
FT CARBOHYD 283 283 POTENTIAL.  
FT CARBOHYD 405 405 POTENTIAL.  
FT CARBOHYD 417 417 POTENTIAL.  
FT CARBOHYD 474 474 POTENTIAL.  
FT CARBOHYD 498 498 POTENTIAL.  
FT CARBOHYD 706 706 POTENTIAL.  
SQ SEQUENCE 931 AA; 102129 MW; 35DB95D7 CRC32;  
Query Match 72.9%; Score 51; DB 1; Length 931;  
Best Local Similarity 55.6%; Pred. No. 2.42e+00;  
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
Db 254 FSLPTFQIL 262  
QY 1 FAMPNFQTL 9  
RESULT 11  
ID YRP2\_RH1ET STANDARD; PRT; 180 AA.  
AC O69777;  
DT 15-DEC-1998 (Rel. 37, Created)  
DT 15-DEC-1998 (Rel. 37, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE HYPOTHETICAL 20.2 KD PROTEIN IN RPN2 3' REGION (ORF180).  
OS Rhizobium etli.  
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
OC Rhizobiaceae; Rhizobium.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-CNPA512;  
RA MICHELIS J., MORIS M., DOMBRECHT B., VERRETH C., VANDERLEYDEN J.;  
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.  
CC -1- SIMILARITY: BELONGS TO THE PEROXIREDOXIN 2 FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC -----  
CC EMBL; Z13965; CAA78367.1; -  
CC PIR; S23853; S23853.  
CC PFAM; PF00850; Hist\_deacetyl; 1.  
CC Hypothetical protein; Hydrolase.  
SQ SEQUENCE 310 AA; 34145 MW; 80821D2D CRC32;  
Query Match 70.0%; Score 49; DB 1; Length 310;  
Best Local Similarity 55.6%; Pred. No. 6.50e+00;  
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;  
Db 31 FPMFKFRL 39  
QY 1 FAMPNFQTL 9  
RESULT 13  
ID THDF\_BUCAP STANDARD; PRT; 456 AA.  
AC Q44633;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 13-DEC-1998 (Rel. 37, Last annotation update)  
DE POSSIBLE THIOPHENE AND FURAN OXIDATION PROTEIN THDF.  
GN THDF.  
OS Buchnera aphidicola.  
OC Bacteria; Proteobacteria; gamma subdivision; Buchnera.  
RN [1]

CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC -----  
CC EMBL; AJ005696; CAA06680.1; -  
CC PFAM; PF00578; Ahpc-TSA; 1.  
CC Hypothetical protein.  
SQ SEQUENCE 180 AA; 20180 MW; 1A5EE847 CRC32;  
Query Match 70.0%; Score 49; DB 1; Length 180;  
Best Local Similarity 44.4%; Pred. No. 6.50e+00;  
Matches 4; Conservative 4; Mismatches 1; Indels 0; Gaps 0;  
Db 60 FOLPDFESL 68  
QY 1 FAMPNFQTL 9  
RESULT 12  
ID YGLA\_SYNP2 STANDARD; PRT; 310 AA.  
AC P28606;  
DT 01-DEC-1992 (Rel. 24, Created)  
DT 01-DEC-1992 (Rel. 24, Last sequence update)  
DT 15-JUL-1998 (Rel. 36, Last annotation update)  
DE HYPOTHETICAL 34.1 KD PROTEIN IN GLNA 3' REGION.  
OS Synechococcus sp. (strain PCC 7002) (Agmenellum quadruplicatum).  
OC Bacteria; Cyanobacteria; Chroococcales; Synechococcus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-PR-6;  
RX MEDLINE; 93139025.  
RA WAGNER S.J., THOMAS S.P., KAUFMAN R.I., NIXON B.T., STEVENS S.E. JR.;  
RT "The glnA gene of the cyanobacterium Agmenellum quadruplicatum PR-6  
RT is nonessential for ammonium assimilation."  
RL J. Bacteriol. 175:604-612(1993).  
CC -1- SIMILARITY: BELONGS TO THE HISTONE DEACETYLASE / ACUC / APHA  
CC FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC -----  
CC EMBL; Z13965; CAA78367.1; -  
CC PIR; S23853; S23853.  
CC PFAM; PF00850; Hist\_deacetyl; 1.  
CC Hypothetical protein; Hydrolase.  
SQ SEQUENCE 310 AA; 34145 MW; 80821D2D CRC32;  
Query Match 70.0%; Score 49; DB 1; Length 310;  
Best Local Similarity 55.6%; Pred. No. 6.50e+00;  
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;  
Db 31 FPMFKFRL 39  
QY 1 FAMPNFQTL 9  
RESULT 13  
ID THDF\_BUCAP STANDARD; PRT; 456 AA.  
AC Q44633;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 13-DEC-1998 (Rel. 37, Last annotation update)  
DE POSSIBLE THIOPHENE AND FURAN OXIDATION PROTEIN THDF.  
GN THDF.  
OS Buchnera aphidicola.  
OC Bacteria; Proteobacteria; gamma subdivision; Buchnera.  
RN [1]

RT "Sequence comparison of woodchuck hepatitis virus replicative forms  
RL shows conservation of the genome."  
RL Virology 162:12-20(1988).  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
CC EMBL; M18752; AAA46767.1; -  
DR PIR; C29969; JDVL7.  
DR PFAM; PF00078; rvt; 1.  
DR PFAM; PF00242; DNA\_pol\_viral\_N; 1.  
DR PFAM; PF00336; DNA\_pol\_viral\_C; 1.  
KW Transferrase; DNA-directed DNA polymerase; DNA replication.  
SQ SEQUENCE 884 AA; 99733 MW; 87604C49 CRC32;  
  
Query Match 72.9%; Score 51; DB 1; Length 884;  
Best Local Similarity 77.8%; Pred. No. 2.42e+00;  
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
  
Db 448 FAVPNLQTL 456  
||:||||  
QY 1 FAMPNFQTL 9  
  
RESULT 9 STANDARD; PRT; 884 AA.  
AC P12899;  
DT 01-OCT-1989 (Rel. 12, Created)  
DT 01-OCT-1989 (Rel. 12, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE DNA POLYMERASE (EC 2.7.7.7).  
GN P.  
OS Woodchuck hepatitis virus 59 (WHV 59).  
OC Viruses; Retroviridae; Hepadnaviridae; Orthohepadnavirus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 88101359.  
RA COHEN J. I., MILLER R. H., ROSENBLUM B., DENNISTON K., GERIN J. L.,  
RA PURCELL R. H.;  
RT "Sequence comparison of woodchuck hepatitis virus replicative forms  
RT shows conservation of the genome."  
RL Virology 162:12-20(1988).  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
CC EMBL; M19183; AAA46763.1; -  
DR PIR; G29969; JDVL59.  
DR PFAM; PF00078; rvt; 1.  
DR PFAM; PF00242; DNA\_pol\_viral\_N; 1.  
DR PFAM; PF00336; DNA\_pol\_viral\_C; 1.  
KW Transferrase; DNA-directed DNA polymerase; DNA replication.  
SQ SEQUENCE 884 AA; 99399 MW; B187E46A CRC32;  
  
Query Match 72.9%; Score 51; DB 1; Length 884;  
Best Local Similarity 77.8%; Pred. No. 2.42e+00;  
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
  
Db 448 FAVPNLQTL 456  
||:||||  
QY 1 FAMPNFQTL 9

RESULT 10  
ID EMRL\_MOUSE STANDARD; PRT; 931 AA.  
AC Q61549;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 15-DEC-1999 (Rel. 39, Last annotation update)  
DE CELL SURFACE GLYCOPROTEIN EMRL PRECURSOR (EMRL HORMONE RECEPTOR)  
DE (CELL SURFACE GLYCOPROTEIN F4/80).  
GN EMRL OR GPF480.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN-BALB/C; TISSUE-PERITONEAL CAVITY;  
RX MEDLINE; 96132946.  
RA MCKNIGHT A. J., MACFARLANE A. J., DRI P., TURLEY L., WILLIS A. C.,  
RA GORDON S.;  
RT "Molecular cloning of F4/80, a murine macrophage-restricted cell  
RT surface glycoprotein with homology to the G-protein-linked  
RT transmembrane 7 hormone receptor family."  
RL J. Biol. Chem. 271:486-489(1996).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 97312684.  
RA LIN H. H., STUBBS L. J., MUCENSKI M. L.;  
RT "Identification and characterization of a seven transmembrane hormone  
RT receptor using differential display."  
RL Genomics 41:301-308(1997).  
CC - FUNCTION: PROBABLY INVOLVED IN CELL ADHESION WITHIN TISSUES  
CC AND RECEPTOR SIGNALING.  
CC - SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.  
CC - TISSUE SPECIFICITY: IN MACROPHAGES; BUT ABSENT FROM THOSE WHICH  
CC ARE LOCALIZED WITHIN T-CELL AREAS OF LYMPH NODES AND SPLEEN.  
CC LOW LEVEL OF EXPRESSION ON BLOOD MONOCYTES.  
CC - SIMILARITY: CONTAINS 7 EGF-LIKE DOMAINS.  
CC - SIMILARITY: BELONGS TO FAMILY 2 OF G-PROTEIN COUPLED RECEPTORS.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
CC EMBL; X93328; CAA63720.1; -  
DR EMBL; U66888; AAC53184.1; -  
DR HSSP; P07204; 1FGD.  
DR GCRDB; GCR\_1309;  
DR MGD; MGI:106912; EMRL.  
DR PROSITE; PS00649; G\_PROTEIN\_RECEP\_F2\_1; FALSE\_NEG.  
DR PROSITE; PS00650; G\_PROTEIN\_RECEP\_F2\_2; 1.  
DR PROSITE; PS001010; ASX\_HYDROXYL; 6.  
DR PROSITE; PS01186; EGF\_2; 1.  
DR PROSITE; PS01187; EGF\_CA; 5.  
DR PFAM; PF00008; EGF; 7.  
KW G-protein coupled receptor; Transmembrane; Receptor; Glycoprotein;  
KW EGF-like domain; Repeat; Signal.  
FT SIGNAL 1 27  
FT CHAIN 28 931  
FT DOMAIN 28 644  
FT TRANSMEM 645 672  
FT DOMAIN 673 679  
FT TRANSMEM 680 701  
FT DOMAIN 702 711  
FT TRANSMEM 712 735  
FT DOMAIN 736 754  
FT TRANSMEM 755 776  
FT DOMAIN 777 792  
FT TRANSMEM 793 821  
FT DOMAIN 822 839  
FT TRANSMEM 840 859  
FT POTENTIAL.  
FT CELL SURFACE GLYCOPROTEIN EMRL.  
FT EXTRACELLULAR (POTENTIAL).  
FT POTENTIAL.  
FT CYTOPLASMIC (POTENTIAL).  
FT POTENTIAL.  
FT EXTRACELLULAR (POTENTIAL).  
FT POTENTIAL.  
FT CYTOPLASMIC (POTENTIAL).  
FT POTENTIAL.  
FT EXTRACELLULAR (POTENTIAL).  
FT POTENTIAL.  
FT CYTOPLASMIC (POTENTIAL).  
FT POTENTIAL.

```

5
RESULT
ID DPOL_HPBGS STANDARD; PRT; 881 AA.
AC P03161;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE DNA POLYMERASE (EC 2.7.7.7) (A PROTEIN).
GN P.
OS Ground squirrel hepatitis virus (GSV).
OC Viruses; Retroviral viruses; Hepadnaviridae; Orthohepadnavirus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 84267998.
RA SEEGER C., GANEM D., VARMUS H.E.;
RT "Nucleotide sequence of an infectious molecularly cloned genome of
ground squirrel hepatitis virus.";
RL J. Virol. 51:367-375(1984).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; K02715; AAA46756.1; -
DR PIR; A00709; JDVLS.
DR PFAM; PF00078; rvt; 1.
DR PFAM; PF00242; DNA_pol_viral_N; 1.
DR PFAM; PF00336; DNA_pol_viral_C; 1.
KW Transferase; DNA-directed DNA polymerase; DNA replication.
SQ SEQUENCE 881 AA; 99976 MW; 2295D041 CRC32;

Query Match 72.9%; Score 51; DB 1; Length 881;
Best Local Similarity 77.8%; Pred. No. 2.42e+00;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 445 FAVPNLQTL 453
||:|||||
QY 1 FAMPNFQTL 9

6
RESULT
ID DPOL_HVH8 STANDARD; PRT; 883 AA.
AC P06275;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE DNA POLYMERASE (EC 2.7.7.7).
GN P.
OS Woodchuck hepatitis virus 8 (WHV 8).
OC Viruses; Retroviral viruses; Hepadnaviridae; Orthohepadnavirus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 86062931.
RA KODAMA K., OGASAWARA N., YOSHIKAWA H., MURAKAMI S.;
RT "Nucleotide sequence of a cloned woodchuck hepatitis virus genome:
evolutional relationship between hepadnaviruses.";
RL J. Virol. 56:978-986(1985).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M11082; AAA19183.1; -
DR PIR; A00708; JDVLC2.
DR PFAM; PF00078; rvt; 1.

7
RESULT
ID DPOL_WHV81 STANDARD; PRT; 884 AA.
AC P17356;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE DNA POLYMERASE (EC 2.7.7.7).
GN P.
OS Woodchuck hepatitis virus 8 (infectious clone) (WHV 8).
OC Viruses; Retroviral viruses; Hepadnaviridae; Orthohepadnavirus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 89184524.
RA GIRONES R., COPE P.J., HORNBUCKLE W.E., TENNANT B.C., GERIN J.L.,
RA PURCELL R.H., MILLER R.H.;
RT "Complete nucleotide sequence of a molecular clone of woodchuck
hepatitis virus that is infectious in the natural host.";
RL Proc. Natl. Acad. Sci. U.S.A. 86:1846-1849(1989).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; J04514; -; NOT_ANNOTATED_CDS.
DR PIR; A32397; JDVLM8.
DR PFAM; PF00078; rvt; 1.
DR PFAM; PF00242; DNA_pol_viral_N; 1.
DR PFAM; PF00336; DNA_pol_viral_C; 1.
KW Transferase; DNA-directed DNA polymerase; DNA replication.
SQ SEQUENCE 884 AA; 99708 MW; 907638E8 CRC32;

Query Match 72.9%; Score 51; DB 1; Length 884;
Best Local Similarity 77.8%; Pred. No. 2.42e+00;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 448 FAVPNLQTL 456
||:|||||
QY 1 FAMPNFQTL 9

8
RESULT
ID DPOL_WHV7 STANDARD; PRT; 884 AA.
AC P12898;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE DNA POLYMERASE (EC 2.7.7.7).
GN P.
OS Woodchuck hepatitis virus 7 (WHV 7).
OC Viruses; Retroviral viruses; Hepadnaviridae; Orthohepadnavirus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 88101359.
RA COHEN J.I., MILLER R.H., ROSENBLUM B., DENNISTON K., GERIN J.L.,
RA PURCELL R.H.;
```

```
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
Db 519 FEMPEDEF 527
| | | | |
QY 1 FAMPNFQTL 9

RESULT 2
ID RS5_METVA STANDARD; PRT; 217 AA.
AC P54045;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE 30S RIBOSOMAL PROTEIN S5P.
GN MJ0475.
OS Methanococcus jannaschii.
OC Archaea: Euryarchaeota; Methanococcales; Methanococcaceae;
OC Methanococcus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-JAL-1 / DSM 2651 / ATCC 43067;
RX MEDLINE; 96337999.
RA BULT C.J., WHITE O., OLSEN G.J., ZHOU L., FLEISCHMANN R.D.,
RA SUTTON G.G., BLAKE J.A., FIZGERALD L.M., CLAYTON R.A., GOCAYNE J.D.,
RA KERLAVAGE A.R., DOUGHERTY B.A., TOMB J.F., ADAMS M.D., REICH C.I.,
RA OVERBECK R., KIRKNESS E.F., WEINSTOCK K.G., MERRICK J.M., GLODER A.,
RA SCOTT J.L., GEOGHAGEN N.S.M., WEIDMAN J.F., FUHRMANN J.L., NGUYEN D.,
RA UTERBACK T.R., KELLEY J.M., PETERSON J.D., SADOW P.W., HANNA M.C.,
RA COTTON M.D., ROBERTS K.M., HURST M.A., KAINE B.P., BORODOVSKY M.,
RA KLENK H.-P., FRASER C.M., SMITH H.O., WOESE C.R., VENTER J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii."
RL Science 273:1058-1073(1996).
CC -!- SIMILARITY: BELONGS TO THE S5P FAMILY OF RIBOSOMAL PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U67497; AAB98464.1; -.
CC HSSP; P02357; 1PKP.
CC TIGR; MJ0475; -.
CC PROSITE; PS00585; RIBOSOMAL_S5; 1.
CC PFAM; PF00333; Ribosomal_S5; 1.
CC KW Ribosomal protein.
CC SEQUENCE 217 AA; 23839 MW; 4D36A4B3 CRC32;

Query Match 72.9%; Score 51; DB 1; Length 217;
Best Local Similarity 55.6%; Pred. No. 2.42e+00;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
Db 185 FAMATFEAL 193
| | | | |
QY 1 FAMPNFQTL 9

RESULT 3
ID DPOL_WHV66 STANDARD; PRT; 556 AA.
AC P11292;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE DNA POLYMERASE (EC 2.7.7.7) (FRAGMENT).
GN P.
OS Woodchuck hepatitis virus w64 (isolate pWS23).
OC Viruses; Retroviridae; Hepadnaviridae; Orthohepadnavirus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 87219879.

Query Match 72.9%; Score 51; DB 1; Length 217;
Best Local Similarity 55.6%; Pred. No. 2.42e+00;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
Db 185 FAMATFEAL 193
| | | | |
QY 1 FAMPNFQTL 9

RESULT 4
ID DPOL_WHV1 STANDARD; PRT; 879 AA.
AC P03160;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE DNA POLYMERASE (EC 2.7.7.7).
GN P.
OS Woodchuck hepatitis virus 1 (WHV 1).
OC Viruses; Retroviridae; Hepadnaviridae; Orthohepadnavirus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 82216969.
RA GALIBERT F., CHEN T.N., MANDART E.;
RT "Nucleotide sequence of a cloned woodchuck hepatitis virus genome:
RT comparison with the hepatitis B virus sequence."
RL J. Virol. 41:51-65(1982).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; J02442; AAA46759.1; -.
CC PIR; A00707; JDLVLC.
CC PFAM; PF00078; rvt; 1.
CC PFAM; PF00242; DNA_pol_viral_N; 1.
CC PFAM; PF00336; DNA_pol_viral_C; 1.
CC KW Transferase; DNA-directed DNA polymerase; DNA replication.
CC SEQUENCE 879 AA; 99185 MW; 3BD450AF CRC32;

Query Match 72.9%; Score 51; DB 1; Length 879;
Best Local Similarity 77.8%; Pred. No. 2.42e+00;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
Db 443 FAVPNLQTL 451
| | | | |
QY 1 FAMPNFQTL 9

RA ETIEMBLE J., MOEROEY T., TREPO C., TIOLLAIS P., BUENDIA M.-A.;
RT "Nucleotide sequence of the woodchuck hepatitis virus surface antigen
RT mRNAs and the variability of three overlapping viral genes."
RL Gene 50:207-214(1986).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M15954; AAA69573.1; -.
CC PIR; A29498; JDLVLC.
CC DR PIR; A29498; rvt; 1.
CC DR PFAM; PF00078; rvt; 1.
CC DR PFAM; PF00242; DNA_pol_viral_N; 1.
CC DR PFAM; PF00336; DNA_pol_viral_C; 1.
CC KW Transferase; DNA-directed DNA polymerase; DNA replication.
CC FT NON_TER 1
CC SQ SEQUENCE 556 AA; 61871 MW; D64F0695 CRC32;

Query Match 72.9%; Score 51; DB 1; Length 556;
Best Local Similarity 77.8%; Pred. No. 2.42e+00;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
Db 120 FAVPNLQTL 128
| | | | |
QY 1 FAMPNFQTL 9
```

\*\*\*\*\*  
M O S E R H  
(TM)  
\*\*\*\*\*  
Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Fri Apr 14 23:25:29 2000; Maspar time 6.14 Seconds  
43.745 Million cell updates/sec  
Tabular output not generated.

Title: >US-08-452-843-5  
Description: (1-9) from US08452843.pep  
Perfect Score: 70  
Sequence: 1 FAMPNFQTL 9  
Scoring table: PAM 150  
Gap 15  
Searched: 82229 seqs, 29864866 residues  
Post-processing: Minimum Match 0%  
Listing first 45 summaries  
Database: swiss-prot38  
1:swissprot  
Statistics: Mean 24.145; Variance 27.237; scale 0.886

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	53	75.7	744	1	BISC_RHOSH	BIOTIN SULFOXIDE REDUC
2	51	72.9	217	1	RS5_METJA	30S RIBOSOMAL PROTEIN
3	51	72.9	556	1	DPOL_RHVW6	DNA POLYMERASE (EC 2.7
4	51	72.9	879	1	DPOL_RHV1	DNA POLYMERASE (EC 2.7
5	51	72.9	881	1	DPOL_RHVB8	DNA POLYMERASE (EC 2.7
6	51	72.9	883	1	DPOL_RHV8	DNA POLYMERASE (EC 2.7
7	51	72.9	884	1	DPOL_RHV81	DNA POLYMERASE (EC 2.7
8	51	72.9	884	1	DPOL_RHV7	DNA POLYMERASE (EC 2.7
9	51	72.9	884	1	DPOL_RHV59	DNA POLYMERASE (EC 2.7
10	51	72.9	931	1	EMRI_MOUSE	CELL SURFACE GLYCOPROT
11	49	70.0	180	1	YRPL_RH1ET	HYPOTHETICAL 20.2 KD P
12	49	70.0	310	1	YGLA_SYNP2	HYPOTHETICAL 34.1 KD P
13	49	70.0	456	1	THDF_BUCAP	POSSIBLE THIOPHENE AND
14	49	70.0	767	1	AMP3_HUMAN	AMP DEAMINASE 3 (EC 3.
15	49	70.0	829	1	TORA_SHEMA	TRIMETHYLAMINE-N-OXIDE
16	48	68.6	306	1	YL86_CAEEL	HYPOTHETICAL 34.6 KD P
17	48	68.6	481	1	DPOL_RHVBW	DNA POLYMERASE (EC 2.7
18	48	68.6	563	1	HEMA_TAMAA	HEMA_TAMAA
19	48	68.6	564	1	HEMA_IACKA	HEMA_TAMAA
20	48	68.6	564	1	HEMA_IADAL	HEMA_TAMAA
21	48	68.6	564	1	HEMA_IABUD	HEMA_TAMAA
22	48	68.6	564	1	HEMA_IATKM	HEMA_TAMAA
23	48	68.6	564	1	HEMA_IADNZ	HEMA_TAMAA

24	48	68.6	564	1	HEMA_IASE2	HEMAAGGLUTININ PRECURSO
25	48	68.6	568	1	HEMA_IAMAB	HEMAAGGLUTININ PRECURSO
26	48	68.6	657	1	HCYA_PANIN	HEMOCYANIN B CHAIN.
27	48	68.6	657	1	HCYA_PANIN	HEMOCYANIN A CHAIN.
28	48	68.6	730	1	DPOL_RHVB4	DNA POLYMERASE (EC 2.7
29	48	68.6	750	1	DPOL_RHVB2	DNA POLYMERASE (EC 2.7
30	48	68.6	763	1	DPOL_RHBPV	DNA POLYMERASE (EC 2.7
31	48	68.6	832	1	DPOL_RHBPV	DNA POLYMERASE (EC 2.7
32	48	68.6	832	1	DPOL_RHBPV	DNA POLYMERASE (EC 2.7
33	48	68.6	832	1	DPOL_RHBPV	DNA POLYMERASE (EC 2.7
34	48	68.6	832	1	DPOL_RHBPV	DNA POLYMERASE (EC 2.7
35	48	68.6	842	1	DPOL_RHBPV	DNA POLYMERASE (EC 2.7
36	48	68.6	843	1	DPOL_RHBPV	DNA POLYMERASE (EC 2.7
37	48	68.6	843	1	DPOL_RHBPV	DNA POLYMERASE (EC 2.7
38	48	68.6	843	1	DPOL_RHBPV	DNA POLYMERASE (EC 2.7
39	48	68.6	843	1	DPOL_RHBPV	DNA POLYMERASE (EC 2.7
40	48	68.6	843	1	DPOL_RHBPV	DNA POLYMERASE (EC 2.7
41	48	68.6	845	1	DPOL_RHBPV	DNA POLYMERASE (EC 2.7
42	48	68.6	845	1	DPOL_RHBPV	DNA POLYMERASE (EC 2.7
43	48	68.6	959	1	MML4_MYCTU	PUTATIVE MEMBRANE PROT
44	48	68.6	967	1	MML4_MYCTU	PUTATIVE MEMBRANE PROT
45	48	68.6	2476	1	ZAN_PIG	ZONADHESIN PRECURSOR.

ALIGNMENTS

RESULT 1  
ID BISC\_RHOSH STANDARD; PRT; 744 AA.  
AC P54934;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE BIOTIN SULFOXIDE REDUCTASE (EC 1.-.-.-) (BDS REDUCTASE) (BSO REDUCTASE)  
OS Rhodospirillum rubrum (Rhodospirillum rubrum)  
OC Bacteria; Proteobacteria; alpha subdivision; Rhodospirillum rubrum  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-SP. DENITRIFICANS IL106;  
RX MEDLINE; 95251380.  
RA POLLOCK V.V., BARBER M.J.;  
RT "Molecular cloning and expression of biotin sulfoxide reductase from Rhodospirillum rubrum forma sp. denitrificans.";  
RL Arch. Biochem. Biophys. 318:322-332(1995).  
CC -1- FUNCTION: THIS ENZYME MAY SERVE AS A SCAVENGER, ALLOWING THE CELL TO UTILIZE BIOTIN SULFOXIDE AS A BIOTIN SOURCE (BY SIMILARITY).  
CC -1- CATALYTIC ACTIVITY: REDUCES A SPONTANEOUS OXIDATION PRODUCT OF BIOTIN, D-BIOTIN D-SULFOXIDE (BSO OR BDS), BACK TO BIOTIN.  
CC -1- COFACTOR: MOLYBDENUM (MOLYBDOTERIN).  
CC -1- SIMILARITY: BELONGS TO THE PROKARYOTIC MOLYBDOTERIN-CONTAINING OXIDOREDUCTASE FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).  
CC EMBL; U08189; AAA74739.1;  
CC HSP; Q57366; 1CXT.  
DR PROSITE; PS00551; MOLYBDOTERIN\_PROK\_1; FALSE\_NEG.  
DR PROSITE; PS00490; MOLYBDOTERIN\_PROK\_2; 1.  
DR PROSITE; PS00932; MOLYBDOTERIN\_PROK\_3; FALSE\_NEG.  
DR PFAM; PF00384; molybdopterin; 1.  
DR PFAM; PF01568; Molybdop\_binding; 1.  
KW Oxidoreductase; Molybdenum.  
SQ SEQUENCE 744 AA; 80266 MW; 6B6E3E56 CRC32;

Query Match 75.7%; Score 53; DB 1; Length 744;  
Best Local Similarity 55.6%; Pred. No. 8.78e-01;



THIS PAGE BLANK (uspto)

DR FLYBASE; FBgn0004367; mei-41.  
DR PFAM; PF00454; P13\_P14\_kinase; 1.  
SQ SEQUENCE 2354 AA; 270445 MW; 1E0FBA72 CRC32;  
Query Match 71.4%; Score 50; DB 5; Length 2354;  
Best Local Similarity 55.6%; Pred. No. 1.12e+01;  
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
Db 764 FVMSRFQSL 772  
| :|:|:|  
Qy 1 FAMPNFQTL 9  
RESULT 15  
ID 086070 PRELIMINARY; PRT; 112 AA.  
AC 086070;  
DT 01-NOV-1998 (TRENBLrel. 08, Created)  
DT 01-NOV-1998 (TRENBLrel. 08, Last sequence update)  
DT 01-NOV-1998 (TRENBLrel. 08, Last annotation update)  
DE HYPOTHETICAL 12.8 KD PROTEIN (FRAGMENT).  
OS Rhizobium etli.  
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
OC Rhizobiaceae; Rhizobium.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CE-3;  
RA SOBERON M., MOREIRA C., MIRANDA-RIOS J., KONDOROSI A., LOPEZ O.;  
RT "A purine related metabolite negatively regulates fixNOOP expression  
in Sinorhizobium meliloti by the modulation of FixK and FnrN  
transcriptional activities.";  
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF083917; AAC34466.1; -.  
KW Hypothetical protein.  
FT NON\_TER 112  
SQ SEQUENCE 112 AA; 12757 MW; 581D55A4 CRC32;  
Query Match 70.0%; Score 49; DB 2; Length 112;  
Best Local Similarity 44.4%; Pred. No. 1.77e+01;  
Matches 4; Conservative 4; Mismatches 1; Indels 0; Gaps 0;  
Db 59 FQLPDFESL 67  
| :|:|:|  
Qy 1 FAMPNFQTL 9  
Search completed: Fri Apr 14 23:28:16 2000  
Job time : 105 secs.

OC Viruses; sRNA positive-strand viruses, no DNA stage; Furovirus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-AHLUM;  
RX MEDLINE; 97170767.  
RA KOENIG R., COMMANDEUR U., LOSS S., BEIER C., KAUFMANN A.,  
RA LESEMAN D.E.;  
RT "Beet soil-borne virus RNA 2: similarities and dissimilarities to the  
RT coat protein gene-carrying RNAs of other furoviruses.";  
RL J. Gen. Virol. 78:469-477(1997).  
DR EMBL; U64512; AAB47479.1; -;  
SQ SEQUENCE 931 AA; 103802 MW; 92A51B31 CRC32;  
  
Query Match 72.9%; Score 51; DB 14; Length 931;  
Best Local Similarity 66.7%; Pred. No. 6.98e+00;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
  
Db 586 FAMPATISL 594  
|||||:  
QY 1 FAMPNFOTL 9  
  
RESULT 11  
ID O68592 PRELIMINARY; PRT; 249 AA.  
AC O68592;  
DT 01-AUG-1998 (TRENBLrel. 07, Created)  
DT 01-AUG-1998 (TRENBLrel. 07, Last sequence update)  
DT 01-MAY-1999 (TRENBLrel. 10, Last annotation update)  
DE STRESS FACTOR A.  
GN PSFA.  
OS Pseudomonas aeruginosa.  
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonas group;  
OC Pseudomonas.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-PA01;  
RX MEDLINE; 96210657.  
RA OCHSNER U.A., VASIL M.L.;  
RT "Gene repression by the ferric uptake regulator in Pseudomonas  
RT aeruginosa: cycle selection of iron-regulated genes.";  
RL Proc. Natl. Acad. Sci. U.S.A. 93:4409-4414(1996).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-PA01;  
RA OCHSNER U.A., VASIL A.I., JOHNSON Z., VASIL M.L.;  
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF051691; AAC06217.1; -;  
DR PFAM; PF00043; GST; 1.  
SQ SEQUENCE 249 AA; 28270 MW; 7990BD28 CRC32;  
  
Query Match 71.4%; Score 50; DB 2; Length 249;  
Best Local Similarity 66.7%; Pred. No. 1.12e+01;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
  
Db 198 FALPAQHL 206  
|||||:  
QY 1 FAMPNFOTL 9  
  
RESULT 12  
ID O04459 PRELIMINARY; PRT; 449 AA.  
AC O04459;  
DT 01-JUL-1997 (TRENBLrel. 04, Created)  
DT 01-JUL-1997 (TRENBLrel. 04, Last sequence update)  
DT 01-NOV-1999 (TRENBLrel. 12, Last annotation update)  
DE F21J9.21.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;  
OC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;  
OC Arabidopsis.  
RN [1]  
RP SEQUENCE FROM N.A.

RA DEWAR K., BUEHLER E., FENG J., KIM C., LI Y., SHINN P., SUN H.,  
RA CONWAY A., CONWAY A., KURTZ D., OJI O., OSBORNE B., SHEN Y.K.,  
RA TORIUMI M., VYSOTSKAIA V., YU G., DAVIS R.W., FEDERSPIEL N.A.,  
RA THEOLOGIS A., ECKER J.R.;  
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AC000103; AAB61524.1; -;  
DR MENDEL; 16595; Arath; 2546; 16595.  
DR PFAM; PF01490; Aa\_trans; 1.  
SQ SEQUENCE 449 AA; 50062 MW; E4E93445 CRC32;  
  
Query Match 71.4%; Score 50; DB 10; Length 449;  
Best Local Similarity 66.7%; Pred. No. 1.12e+01;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
  
Db 328 FAMPVFDML 336  
|||||:  
QY 1 FAMPNFOTL 9  
  
RESULT 13  
ID Q9Y526 PRELIMINARY; PRT; 1181 AA.  
AC Q9Y526;  
DT 01-NOV-1999 (TRENBLrel. 12, Created)  
DT 01-NOV-1999 (TRENBLrel. 12, Last sequence update)  
DT 01-NOV-1999 (TRENBLrel. 12, Last annotation update)  
DE DJ439F8.2 (NOVEL KIAA0279 LIKE CADHERIN DOMAIN PROTEIN (SIMILAR TO  
DE MOUSE CELSR1, RAT MEGF2)) (FRAGMENT).  
GN DJ439F8.2.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA WILLIAMS S.;  
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.  
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).  
DR EMBL; AL021392; CAB50707.1; -;  
DR PROSITE; PS00232; CADHERIN; 7.  
KW Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat.  
FT NON\_TER 1181  
SQ SEQUENCE 1181 AA; 128318 MW; A1567A1D CRC32;  
  
Query Match 71.4%; Score 50; DB 4; Length 1181;  
Best Local Similarity 71.4%; Pred. No. 1.12e+01;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
  
Db 245 FPMNPVQ 251  
|-|||:  
QY 1 FAMPNFQ 7  
  
RESULT 14  
ID Q24135 PRELIMINARY; PRT; 2354 AA.  
AC Q24135;  
DT 01-NOV-1996 (TRENBLrel. 01, Created)  
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)  
DT 01-NOV-1998 (TRENBLrel. 08, Last annotation update)  
DE MEI-41.  
GN MEI-41.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Insecta; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-1-54.2; 14C4-6;  
RX MEDLINE; 95401271.  
RA HAWLEY R.S.;  
RA HARI K.L., SENTERRE A., SEKELSKY J.J., MCKIM K.S., BOYD J.B.,  
RT "The mei-41 gene of D. melanogaster is a structural and functional  
RT homolog of the human ataxia telangiectasia gene.";  
RL Cell 82:815-821(1995).  
DR EMBL; U34925; AAC46881.1; -;

```
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 97312684.
RA LIN H.H., STUBBS L.J., MUCENSKI M.L.;
RT "Identification and characterization of a seven transmembrane hormone
receptor using differential display.";
RL Genomics 41:301-308(1997).
DR EMBL; U66892; AAC3188.1; -.
DR HSSP; P07204; 1EGT.
DR MGD; MGI:106912; Emr1.
DR PROSITE; PS00010; ASX_HYDROXYL; 4.
DR PROSITE; PS01187; EGF_CA; 4.
DR PFAM; PF00008; EGF; 4.
KW Glycoprotein; EGF-like domain.
FT NON_TER 255
SQ SEQUENCE 255 AA; 27764 MW; 0FA23F4D CRC32;

Query Match 72.9%; Score 51; DB 11; Length 255;
Best Local Similarity 55.6%; Pred. No. 6.98e+00;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 113 FSLPTFOIL 121
|::|::|
QY 1 FAMPNFOTL 9

RESULT 7
ID O08744 PRELIMINARY; PRT; 304 AA.
AC O08744;
DT 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE EGF-LIKE MODULE CONTAINING, MUCIN-LIKE, HORMONE RECEPTOR-LIKE
SEQUENCE 1 (EMR1) (FRAGMENT).
GN EMR1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 97312684.
RA LIN H.H., STUBBS L.J., MUCENSKI M.L.;
RT "Identification and characterization of a seven transmembrane hormone
receptor using differential display.";
RL Genomics 41:301-308(1997).
DR EMBL; U66891; AAC3187.1; -.
DR HSSP; P07204; 1EGT.
DR MGD; MGI:106912; Emr1.
DR PROSITE; PS00010; ASX_HYDROXYL; 5.
DR PROSITE; PS01187; EGF_CA; 4.
DR PFAM; PF00008; EGF; 5.
KW Glycoprotein; EGF-like domain.
FT NON_TER 304
SQ SEQUENCE 304 AA; 33013 MW; F967B4BF CRC32;

Query Match 72.9%; Score 51; DB 11; Length 304;
Best Local Similarity 55.6%; Pred. No. 6.98e+00;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 162 FSLPTFOIL 170
|::|::|
QY 1 FAMPNFOTL 9

RESULT 8
ID Q89244 PRELIMINARY; PRT; 585 AA.
AC Q89244;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAY-1999 (TrEMBLrel. 10, Last annotation update)

DE beet soil-borne virus.

DE POLYMERASE PROTEIN (FRAGMENT).
OS Woodchuck hepatitis virus.
OC Viruses; Retroid viruses; Hepadnaviridae; Orthohepadnavirus.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=TOMPKINS COUNTY, N.Y.;
RX MEDLINE; 93255897.
RA KEW M.C., TENNANT B.C., PURCELL R.H., MILLER R.H.;
RT "Heterogeneity of the woodchuck hepatitis virus genome in a
chronically infected woodchuck.";
RL Virus Res 27:229-237(1993).
DR EMBL; M90520; AAA46774.1; -.
DR PFAM; PF00336; DNA_pol_viral_C; 1.
DR PFAM; PF00242; DNA_pol_viral_N; 1.
DR PFAM; PF00078; rvt; 1.
FT NON_TER 1
SQ SEQUENCE 585 AA; 65216 MW; FFE6E996 CRC32;

Query Match 72.9%; Score 51; DB 14; Length 585;
Best Local Similarity 77.8%; Pred. No. 6.98e+00;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 149 FAVPNLOTL 157
|::|::|
QY 1 FAMPNFOTL 9

RESULT 9
ID Q58430 PRELIMINARY; PRT; 625 AA.
AC Q58430;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE 625AA LONG HYPOTHETICAL THREONYL-TRNA SYNTHETASE.
GN PH0699.
OS Pyrococcus horikoshii.
OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=OT3;
RX MEDLINE; 98344137.
RA KAWARABAYASI Y., SAWADA M., HORIKAWA H., HAIKAWA Y., HINO Y.,
RA YAMAMOTO S., SEKINE M., BABA S., KOSUGI H., HOSOFUKA A., NAGAI Y.,
RA SAKAI M., OGURA K., OTUKA R., NAKAZAWA H., TAKAMIYA M., OHFUKU Y.,
RA FUNAHASHI T., TANAKA T., KUDOH Y., YAMAZAKI J., KUSHIDA N., OGUCHI A.,
RA AOKI K., NAKAMURA Y., ROBB T.F., HORIKOSHI K., MASUCHI Y., SHIZUYA H.,
RA KIKUCHI H.;
RT "Complete sequence and gene organization of the genome of a hyper-
thermophilic archaeobacterium, Pyrococcus horikoshii OT3.";
RL DNA Res. 5:55-76(1998).
DR EMBL; AP000003; BAA29790.1; -.
DR PFAM; PF00587; trna-synt_2b; 1.
DR PRINTS; PR01047; TRNASYNTHTR.
KW Aminocyl-trna synthetase.
FT NON_TER 625
SQ SEQUENCE 625 AA; 73023 MW; 768AFAD9 CRC32;

Query Match 72.9%; Score 51; DB 1; Length 625;
Best Local Similarity 55.6%; Pred. No. 6.98e+00;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 344 FTMPDMHTL 352
|::|::|
QY 1 FAMPNFOTL 9

RESULT 10
ID P87544 PRELIMINARY; PRT; 931 AA.
AC P87544;
DT 01-MAY-1997 (TrEMBLrel. 03, Created)
DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE 104K PROTEIN.
```

RA WILSON R., AINSOUGH R., ANDERSON K., BAYNES C., BERKS M.,  
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,  
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,  
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,  
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,  
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,  
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,  
RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,  
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,  
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.,  
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
RT elegans.";  
RL Nature 368:32-38(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RA DANTE M., KRAMER J., TWYMAN B.;  
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RA WATERSTON R.;  
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AF016666; AAB6095.1; -;  
DR HSSP: P56682; ICCV.  
SQ SEQUENCE 490 AA; 56298 MW; 823A74E0 CRC32;

Query Match 77.1%; Score 54; DB 5; Length 490;  
Best Local Similarity 66.7%; Pred.No. 1.65e+00;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 359 FLMPKFQVL 367  
| | | | |  
QY 1 FAMPNFQTL 9

RESULT 3  
ID Q49735 PRELIMINARY; PRT; 87 AA.  
AC Q49735;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
DT 01-FEB-1997 (TREMBLrel. 02, Last annotation update)  
DE HYPOTHETICAL 9.3 KD PROTEIN B1620\_F1\_14.  
GN B1620\_F1\_14.  
OS Mycobacterium leprae.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA SMITH D.R., ROBISON K.;  
RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.  
DR EMBL: U00015; AAC43239.1; -;  
KW Hypothetical protein.  
SQ SEQUENCE 87 AA; 9272 MW; 6AF56082 CRC32;

Query Match 74.3%; Score 52; DB 2; Length 87;  
Best Local Similarity 55.6%; Pred.No. 4.34e+00;  
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 19 FGMTNFQAM 27  
| | | | |  
QY 1 FAMPNFQTL 9

RESULT 4  
ID Q9X7K5 PRELIMINARY; PRT; 245 AA.  
AC Q9X7K5;  
DT 01-NOV-1999 (TREMBLrel. 12, Created)  
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)  
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)  
DE PUTATIVE NITROGEN FIXATION PROTEIN.  
GN NIFQ.  
OS Rhizobium galegae.

OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
OC Rhizobiaceae; Rhizobium.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA SUOMINEN L., PAULIN L., ROOS C., SAANO A., SAREN A.M., TAS E.,  
RA LINDSTROM K.;  
RT "Identification of nodulation promoter (nod-box) regions of Rhizobium  
RT galegae.";  
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AJ132912; CAB40565.1; -;  
SQ SEQUENCE 245 AA; 27171 MW; 24F32A8B CRC32;

Query Match 74.3%; Score 52; DB 2; Length 245;  
Best Local Similarity 87.5%; Pred.No. 4.34e+00;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 93 FAMPNLOT 100  
| | | | |  
QY 1 FAMPNFQTL 8

RESULT 5  
ID Q22215 PRELIMINARY; PRT; 262 AA.  
AC Q22215;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
DT 01-JAN-1999 (TREMBLrel. 09, Last annotation update)  
DE T05B9.1 PROTEIN.  
GN T05B9.1.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;  
OC Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA SWINBURNE J.;  
RL Submitted (APR-1995) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 94150718.  
RA WILSON R., AINSOUGH R., ANDERSON K., BAYNES C., BERKS M.,  
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,  
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,  
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,  
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,  
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,  
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,  
RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,  
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,  
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.,  
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
RT elegans.";  
RL Nature 368:32-38(1994).  
DR EMBL: 249129; CAA88962.1; -;  
SQ SEQUENCE 262 AA; 30162 MW; D3F796CF CRC32;

Query Match 74.3%; Score 52; DB 5; Length 262;  
Best Local Similarity 55.6%; Pred.No. 4.34e+00;  
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 79 FQMAHFQSL 87  
| | | | |  
QY 1 FAMPNFQTL 9

RESULT 6  
ID C08745 PRELIMINARY; PRT; 255 AA.  
AC C08745;  
DT 01-JUL-1997 (TREMBLrel. 04, Created)  
DT 01-JUL-1997 (TREMBLrel. 04, Last sequence update)  
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)  
DE EGF-LIKE MODULE CONTAINING, MUCIN-LIKE, HORMONE RECEPTOR-LIKE  
DE SEQUENCE 1 (EMRL) (FRAGMENT).  
GN EMRL.

\*\*\*\*\*

(TM)

\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

protein - protein database search, using Smith-Waterman algorithm

```
Run on:      Fri Apr 14 23:26:31 2000;  MasPar time 12.80 Seconds
            48.765 Million cell updates/sec
```

Tabular output not generated.

```

Title: >US-08-452-843-5
Description: (1-9) from US08452843.pep
Perfect Score: 70
Sequence: 1 FAMPNFQTL 9

```

Scoring table: PAM 150  
Gap 15

Searched: 225878 seqs, 69334122 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: sptrembl12  
1:sp\_archaea 2:sp\_bacteria 3:sp\_fungi 4:sp\_human  
5:sp\_invertebrate 6:sp\_mammal 7:sp\_mhc 8:sp\_organelle  
9:sp\_phase 10:sp\_plant 11:sp\_rodent 12:sp\_unclassified  
13:sp\_vertebrate 14:sp\_virus

Statistics: Mean 23.047; Variance 28.729; scale 0.802

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Query Match	Score	Length	DB ID	Description	Pred. No.
1	62	88.6	952 14	Q36413	TEGUMENT PROTEIN.	2.79e+02
2	54	77.1	490 5	Q16620	F35H9.4 PROTEIN.	1.65e+00
3	52	74.3	87 2	Q49735	HYPOTHETICAL 9.3 KD PR	4.34e+00
4	52	74.3	245 2	Q9X7K5	PUTATIVE NITROGEN FIXA	4.34e+00
5	52	74.3	262 5	Q22215	T05B9.1 PROTEIN.	4.34e+00
6	51	72.9	255 11	Q08745	EGF-LIKE MODULE CONTAI	6.98e+00
7	51	72.9	304 11	Q08744	EGF-LIKE MODULE CONTAI	6.98e+00
8	51	72.9	585 14	Q89244	POLYMERASE PROTEIN (FR	6.98e+00
9	51	72.9	625 1	Q58430	625AA LONG HYPOTHETICA	6.98e+00
10	51	72.9	931 14	Q87544	104K PROTEIN.	6.98e+00
11	50	71.4	249 2	Q68592	STRESS FACTOR A.	1.12e+01
12	50	71.4	449 10	Q04459	F21J9.21.	1.12e+01
13	50	71.4	1181 4	Q9Y526	DJ439F8.2 (NOVEL KIAA0	1.12e+01
14	50	71.4	2354 5	Q24135	MEI-41.	1.12e+01
15	49	70.0	112 2	Q68070	HYPOTHETICAL 12.8 KD P-	1.12e+01
16	49	70.0	180 5	Q22587	COSMID T19D7.	1.77e+01
17	49	70.0	186 2	Q06450	DNAG, RPOD, CPOA GENES	1.77e+01
18	49	70.0	300 2	Q44173	HYPOTHETICAL 33.0 KD P	1.77e+01
19	48	68.6	730 14	Q9XKJ8	HYPOTHETICAL 39.3K D	2.80e+01
20	48	68.6	801 14	Q9WP64	POLYMERASE.	2.80e+01
21	48	68.6	730 14	Q9WP64	P PROTEIN.	2.80e+01

## ALIGNMENTS

RESULT	1
ID	O36413
AC	O36413; PRELIMINARY; PRT; 952 AA.
DT	01-JAN-1998 (TrEMBLrel. 05, Created)
DT	01-JUN-1998 (TrEMBLrel. 05, Last sequence update)
DT	01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE	TGEMENT PROTEIN.
OS	Alcalaphine herpesvirus 1 (wildbeest herpesvirus).
OC	Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
CC	Gammaherpesvirinae.
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN-C500;
RX	MEDLINE; 97404559.
RA	ENSSLER A., PFLANZ R., FLECKENSTEIN B.;
RL	"Primary structure of the alcalaphine herpesvirus 1 genome.";
RT	J. Virol. 71:6517-6525(1997).
DR	EBML; AF005370; AAC58110.1; -.
SD	SEQUENCE 952 AA; 107065 MW; 0239B40F CRC32:

Query Match 88.6%; Score 62; DB 14; Length 952;  
Best Local Similarity 77.8%; Pred. No. 2.79e-02

DB 185 FNMPNFQTM 193  
QY 1 FAMPNFQTL 9

RESULT	2	PRELIMINARY;	PRT:	490 AA.
ID	Q16620			
AC	Q16620			
DT	01-JAN-1998	(TREMBLrel. 05, Created)		
DT	01-JAN-1998	(TREMBLrel. 05, Last sequence up		
DT	01-NOV-1999	(TREMBLrel. 12, Last annotation		
DE	F36H9.4 PROTEIN.			

```

FT peptide 637..748
FT /note="peptide of claim 1"
FT peptide 738..745
FT /note="peptide of claim 1"
PN J08023972-A.
PD 30-JAN-1996.
PF 19-JUL-1994; 187936.
PR 19-JUL-1994; JP-187936.
PA (SUNR ) SUNTORY LTD.
DR WPI; 96-133414/14.
DR N-PSDB; T11575.
PT New glucagon decomposing enzyme, and DNA encoding it - for
PT specifically cleaving glucagon and vasoactive intestinal peptide, in
PT the prevention and treatment of diseases caused by excess glucagon
PT and Vip
PS Claim 2; Page 2; 18pp; Japanese.
CC This is the amino acid sequence of a novel isolated glucagon degrading
CC enzyme (GDE) of mol. wt. 83 kD. The enzyme has a pH optimum of 6.8 and
CC catalyses the cleavage of glucagon, vasoactive intestinal peptide and
CC selectin (R93022-4). The corresp. gene was isolated from a human
CC pancreatic carcinoma cell line HPC-Yo cDNA library by screening the
CC library with an anti-GDE peptide antibody, amplifying the inserts with
CC the primers T18903-4 and probing the fragments with the probe T18905.
CC This screening resulted in the full length clone designated lambda
CC GDE4-2. The coding region of the clone was subsequently PCR amplified by
CC the primers T11576-7 and inserted into the eukaryotic expression vector
CC pKDCR under control of the SV40 promoter for production of the protein in
CC COS-7 cells. The protein is useful in preventing and treating diseases
CC characterised by an excess of glucagon or vasoactive intestinal peptide.
SQ Sequence 864 AA;

Query Match 66.7%; Score 50; DB 1; Length 864;
Best Local Similarity 55.6%; Pred. No. 2.112e+02;
Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Db 412 F5MDYFYGL 420
Qy 1 FAMPNFYTL 9

```

Search completed: Fri Apr 14 23:30:47 2000  
Job time : 42 secs.

CC invention. The polypeptides can be used for preventing or treating  
CC Helicobacter infections, and gastroduodenal diseases associated with  
CC these infections, including acute, chronic, and atrophic gastritis, and  
CC peptic ulcer diseases, e.g. gastric and duodenal ulcers. They can also be  
CC used for the production of antibodies. The products can also be used for  
CC detection and diagnosis.

SQ sequence 381 AA;  
Query Match 66.7%; Score 50; DB 1; Length 381;  
Best Local Similarity 57.1%; Pred. No. 2.12e+02;  
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 159 FGMPNYF 165  
QY 1 FAMPNFI 7

RESULT 13  
ID W55637 standard; Protein; 381 AA.

AC W55637;  
DT 03-JUL-1998 (first entry)  
DE H. pylori ORF 03a10804\_21698400\_c2\_32 cytoplasmic protein.  
KW Cytoplasmic; vaccine; prevention; treatment; infection; envelope;  
KW identification; binding compound; bacteria; life cycle; activator;  
KW inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis.

OS Helicobacter pylori.  
PN W09737044-A1.  
PD 09-OCT-1997;  
PF 27-MAR-1997; U05223.  
PR 06-DEC-1996; US-761318.  
PR 29-MAR-1996; US-625811.  
PR 02-APR-1996; US-758731.  
PR 25-OCT-1996; US-736905.  
PR 28-OCT-1996; US-738859.  
PA (ASTR ) ASTRA AB.  
PI Alm RA, Smith D;  
DR WPI; 97-503122/46.  
DR N-PSDB; V25046.  
PT Helicobacter pylori nucleic acid sequences and encoded  
PT polypeptide(s) - useful in vaccines to treat or prevent H. pylori  
PT infection and for diagnosis of H. pylori infection

PS Claims 14,96; Page 836-837; 1145pp; English.  
CC This sequence is a H. pylori cytoplasmic protein. The protein may be  
CC used in a vaccine to prevent or treat H. pylori infection or to identify  
CC H. pylori polypeptide binding compounds, useful as potential H. pylori  
CC life cycle activators or inhibitors. The DNA and probes derived from it  
CC may be used for the identification of H. pylori in a sample and the  
CC diagnosis of H. pylori infection. Nucleic acid sequences complementary to  
CC the DNA act as antisense sequences and can be used to prevent the  
CC translation of H. pylori mRNA. Antibodies against the protein can be used  
CC in immunoassays to evaluate the abundance and distribution of  
CC H. pylori-specific antigens. The genomic sequence of H. pylori  
CC (ATCC 55679) was determined from overlapping contigs generated by  
CC mechanically shearing the bacterial DNA. The sequences were analysed  
CC for ORF of at least 180 nucleotides, and the predicted coding regions  
CC defined by computer evaluation. To identify likely H. pylori antigens for  
CC vaccine development, the amino acid sequences predicted from various ORF  
CC were analysed for significant homology to other known or exported  
CC membrane proteins. Having identified and determined the sequences of  
CC interest, particular regions can be isolated from H. pylori by PCR  
CC amplification for recombinant polypeptide production, e.g. in E. coli  
CC hosts.

SQ Sequence 381 AA;

Query Match 66.7%; Score 50; DB 1; Length 381;  
Best Local Similarity 57.1%; Pred. No. 2.12e+02;  
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 159 FGMPNYF 165  
QY 1 FAMPNFI 7

RESULT 14

ID W55311 standard; Protein; 386 AA.  
AC W55311;  
DT 15-JUN-1998 (first entry)  
DE H. pylori ORF 06p10306r11 protein.  
KW Cytoplasmic; vaccine; prevention; treatment; infection; envelope;  
KW identification; binding compound; bacteria; life cycle; activator;  
KW inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis.

OS Helicobacter pylori.  
PN W09737044-A1.  
PD 09-OCT-1997;  
PF 27-MAR-1997; U05223.  
PR 06-DEC-1996; US-761318.  
PR 29-MAR-1996; US-625811.  
PR 02-APR-1996; US-758731.  
PR 25-OCT-1996; US-736905.  
PR 28-OCT-1996; US-738859.  
PA (ASTR ) ASTRA AB.  
PI Alm RA, Smith D;  
DR WPI; 97-503122/46.  
DR N-PSDB; V24720.

PT Helicobacter pylori nucleic acid sequences and encoded  
PT polypeptide(s) - useful in vaccines to treat or prevent H. pylori  
PT infection and for diagnosis of H. pylori infection

PS Claim 14; Page 540-541; 1145pp; English.  
CC This sequence is a H. pylori protein of unspecified function.  
CC The protein may be used in a vaccine to prevent or treat H. pylori  
CC infection or to identify H. pylori polypeptide binding compounds,  
CC useful as potential H. pylori life cycle activators or inhibitors. The  
CC DNA and probes derived from it may be used for the identification of  
CC H. pylori in a sample and the diagnosis of H. pylori infection. Nucleic  
CC acid sequences complementary to the DNA act as antisense sequences and  
CC can be used to prevent the translation of H. pylori mRNA. Antibodies  
CC against the protein can be used in immunoassays to evaluate the abundance  
CC and distribution of H. pylori-specific antigens. The genomic sequence of  
CC H. pylori (ATCC 55679) was determined from overlapping contigs generated  
CC by mechanically shearing the bacterial DNA. The sequences were analysed  
CC for ORF of at least 180 nucleotides, and the predicted coding regions  
CC defined by computer evaluation. To identify likely H. pylori antigens for  
CC vaccine development, the amino acid sequences predicted from various ORF  
CC were analysed for significant homology to other known or exported  
CC membrane proteins. Having identified and determined the sequences of  
CC interest, particular regions can be isolated from H. pylori by PCR  
CC amplification for recombinant polypeptide production, e.g. in E. coli  
CC hosts.

SQ Sequence 386 AA;

Query Match 66.7%; Score 50; DB 1; Length 386;  
Best Local Similarity 57.1%; Pred. No. 2.12e+02;  
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 164 FGMPNYF 170  
QY 1 FAMPNFI 7

RESULT 15

ID R93021 standard; Protein; 864 AA.  
AC R93021; 1996 (first entry)  
DE Human glucagon degrading enzyme.  
KW Glucagon degrading enzyme; catalyst; cleavage; selectin; human; primer;  
KW vasoactive intestinal peptide; Vip; pancreatic carcinoma cell line; PCR;  
KW amplification; polymerase chain reaction; probe; expression vector;  
KW eukaryote; SV40 promoter; COS-7.

OS Homo sapiens.  
PI Key Location/Qualifiers  
FT peptide 206. .224  
FT peptide /note= "peptide of claim 1"  
FT peptide 422. .436  
FT peptide /note= "peptide of claim 1"  
FT peptide 528. .536  
FT peptide /note= "peptide of claim 1"



||||| 1:11  
1 FAMPNFYTL 9

QY

RESULT 11

ID W56296 standard; Protein; 245 AA.  
AC W56296;  
DT 28-SEP-1998 (first entry)  
DE Babesia microti BMNI-16 antigen sequence.  
KW antigen; detection; diagnosis; vaccine; tick-borne disease;  
OS differentiation; Lyme disease; ehrlichiosis.  
KW Babesia microti.  
PN EP-834567-A2.  
PD 08-APR-1998.  
PF 01-OCT-1997; 117067.  
PR 24-APR-1997; US-845258.  
PR 01-OCT-1996; US-723142.  
PR (CORI-) CORIXA CORP.  
PA Houghton R, Lodes MJ, Reed SG, Sleath PR;  
PI WPI; 98-195465/18.  
DR N-PSDB: V22746.  
DR

PT vectors, transformed cells and antibodies, useful for diagnosis of  
PT infection and in protective vaccines  
PS Claim 1; page 66; 113pp; English.

CC one antigenic port.

CC in usual immunoa.

RESULT	10
ID	W56297 standard; Protein; 245 AA.

CC allow rapid differentiation between B. microti infection and  
CC other tick-borne diseases ( Lyme disease and ehrlichiosis) that  
CC have similar symptoms but require different treatments.  
SQ Sequence 245 AA;

Query Match 66.7%; Score 50; DB 1; Length 245;

Best Local Similarity 77.88; Pred. No. 2.12e+02;

Db 231 FAMP-FFTL 238

QY 1 FAMPNFYTL 9

RESULT 12  
ID W98429 standard; Protein; 381 AA.

KW GHPO protein; Helicobacter infection; gastroduodenal disease; gastritis;  
 KW peptic ulcer disease.  
 OS Helicobacter pylori.  
 PN W09843478-A1.

PT. New Isolated Helic

PT for the diagnosis, prevention and treatment of

CC This sequence repr

This sequence represents a *Helicobacter pylori* GHPO protein of the

Db 1802 FSLPKFYLL 1810  
 Qy 1 FAMPNFYTL 9

RESULT 6

ID W13280 standard; Protein; 2818 AA.  
 AC W13280;  
 DT 05-JUN-1997 (first entry)  
 DE Human neurofibromin.  
 KW regulation; ras-cAMP; pathway; mammalian; GAP; ras p21; gene;  
 KW activation; neurofibromatosis; type 1; NF1; somatic; mutation;  
 KW tumour; detection; diagnosis; prognosis; defective; treatment.  
 OS Homo sapiens.  
 FH Key  
 FT domain 1175..1534  
 FT /note= "GTPase activating protein (GAP) related  
 FT domain (GRD)"  
 FT 1389..1391  
 FT region  
 FT /note= "conserved region in GRD"

PN US5605799-A.  
 PD 25-FEB-1997.  
 PF 12-JUL-1990; 551531.  
 PR 12-JUL-1990; US-551531.  
 PR 16-APR-1993; US-047088.  
 PR 28-MAR-1995; US-411389.  
 PA (UTAH ) UNIV UTAH RES FOUND.  
 PI Cawthon RM, Li Y, White RL;  
 PI WPI; 97-153572/14.  
 DR N-PSDB; T46941.

PT Detection of defective ras regulation at the neurofibromatosis type  
 PT 1 gene in tumour - by detecting mutation in specified region of gene  
 PS Claim 1: Columns 17-38; 35pp; English.  
 CC The present sequence is human neurofibromin (hNF), which is  
 CC largely homologous to yeast IRA protein (inhibitory regulators of  
 CC the ras-cAMP pathway) and mammalian GAP (ras p21 GTPase activating  
 CC proteins). The hNF gene is the human neurofibromatosis type 1 (NF1)  
 CC gene, somatic mutations of which in the region spanning nucleotides  
 CC 3809-4888 of the NF1 cDNA, in human tumours, indicates defective  
 CC ras regulation. Therefore a tumour found to contain a somatic  
 CC mutation in the NF1 gene can be treated using ras activity as the  
 CC focus, whereas a tumour not containing such a mutation will require  
 CC other courses of treatment. A tumour containing a somatic mutation  
 CC in the NF1 gene can be treated by inactivating ras p21, also as GAP  
 CC p120 is present, but apparently latent, GAP p120 activation would  
 CC be beneficial and finally inhibition of GDP/GTP exchange would also  
 CC counteract the loss of hNF or hNF GAP related domain activity.  
 SQ Sequence 2818 AA;

Query Match 72.0%; Score 54; DB 1; Length 2818;  
 Best Local Similarity 55.6%; Pred. No. 8.69e+01;  
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 2135 FSLPKFYLL 2143  
 Qy 1 FAMPNFYTL 9

RESULT 7

ID R22268 standard; Protein; 2818 AA.  
 AC R22268;  
 DT 06-MAY-1992 (first entry)  
 DE Nfl gene product.  
 KW von Recklinghausen neurofibromatosis disease; autosomal dominant;  
 KW gene therapy.  
 OS Homo sapiens.  
 PN W09200387-A.  
 PD 09-JAN-1992.  
 PF 28-JUN-1991; U04624.  
 PR 29-JUN-1990; US-547090.  
 PA (UNMI ) UNIV OF MICHIGAN.  
 PI Collins FS, Wallace MR, Marchuk DA, Andersen LB, Gutmann DH;

DR WPI; 92-041568/05.  
 DR N-PSDB; Q20602.  
 PT DNA sequences to von-Recklinghausen neurofibromatosis gene - and  
 PT derived amino acid sequences and probes for screening NF1 in early  
 PT stages of disease  
 PS Claim 25; Page 67; 122pp; English.  
 CC This is the amino acid sequence of the von Recklinghausen neuro-  
 CC fibromatosis (NF1) gene product. It and antibodies raised to it  
 CC can be used in hybridisation and immunological assays to screen for  
 CC the presence of a normal or defective Nf1 gene product. Functional  
 CC assays to measure levels of gene function can also be used for  
 CC diagnosis or to monitor treatment. Patient therapy through  
 CC supplementation with the normal Nf1 product which can be  
 CC produced by recombinant techniques is also possible.  
 SQ Sequence 2818 AA;

Query Match 72.0%; Score 54; DB 1; Length 2818;  
 Best Local Similarity 55.6%; Pred. No. 8.69e+01;  
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 2135 FSLPKFYLL 2143  
 Qy 1 FAMPNFYTL 9

RESULT 8

ID R89327 standard; Protein; 741 AA.  
 AC R89327;  
 DT 08-APR-1996 (first entry)  
 DE Membrane anchor protein.  
 KW LKPI operon; peptidase; periplasmic chaperone protein;  
 KW minor tip-associated protein; tip adhesin protein; cloning;  
 KW Escherichia coli; plasmid pHF1; diagnostic; probe; antibody;  
 KW recombinant vaccine.  
 OS Haemophilus influenzae (serotype 1).  
 PN W09603648-A1.  
 PD 01-FEB-1996.  
 PF 13-JUL-1995; U08789.  
 PR 19-JUL-1994; US-277231.  
 PR 07-JUN-1995; US-477326.  
 PR 07-JUN-1995; US-473750.  
 PA (AMCY ) AMERICAN CYANAMID CO.  
 PA (BACT-) BACTEX INC.  
 PI Brinton CC, Green BA;  
 DR WPI; 96-105910/11.  
 DR N-PSDB; Q99312

PT Haemophilus influenzae 1 LKP pilin genes and proteins - used to  
 PT produce anti-H. influenzae antibodies, used to detect and vaccinate  
 PT against H. influenzae  
 PS Claim 2; Page 45-47; 63pp; English.  
 CC The sequence represents a membrane anchor protein encoded by the  
 CC hlpR gene in the LKPI operon from Haemophilus influenzae serotype-1.  
 CC The operon also encodes integrase, pilin protein (R89325),  
 CC periplasmic chaperone protein (R89326), minor tip-associated protein  
 CC (R89328), tip adhesin protein (R89329) and peptidase. The operon  
 CC has been isolated by cloning in Escherichia coli using plasmid pHF1.  
 CC The operon and its encoded proteins may be used in production of  
 CC diagnostic probes, antibodies and recombinant vaccines.  
 SQ Sequence 741 AA;

Query Match 70.7%; Score 53; DB 1; Length 741;  
 Best Local Similarity 55.6%; Pred. No. 1.09e+02;  
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 368 YSSPDFYTL 376  
 Qy 1 FAMPNFYTL 9

RESULT 9

ID R88469 standard; Protein; 1464 AA.  
 AC R88469;

DE Cw3 consensus peptide derived immunogenic peptide #1.  
 KW Immunogenic peptide; supermotif; HLA molecule; CTL response;  
 KW therapeutic; diagnostic; cancer; viral infection; hepatitis B;  
 KW hepatitis C.  
 OS Synthetic.  
 PN WO9603140-A1.  
 PD 08-FEB-1996.  
 PF 21-JUL-1995; U09234.  
 PR 21-JUL-1994; US-278634.  
 PR 23-NOV-1994; US-344824.  
 PR 30-MAY-1995; US-452843.  
 PA (CYTE-) CYTEL CORP.  
 PI Sette A, Sidney J;  
 DR WPI; 96-116784/12.  
 PT Compn. comprising immunogenic peptide with supermotif allowing more  
 PT than one HLA mol. to bind - used to induce CTL response in patient  
 PT and for in vivo and ex vivo therapeutic and diagnostic applications  
 PS Claim 2; Page 26; 32pp; English.  
 CC The sequences given in R89362-82 are immunogenic peptides which were  
 CC use in the composition of the invention. The composition comprises  
 CC an immunogenic peptide of 9-10 residues with a supermotif which  
 CC allows binding of more than one HLA molecule. It pref. comprises  
 CC two conserved residues, a first at the 2nd position from the N-  
 CC terminal is Pro, and a 2nd at the C-terminal is Met. These peptides  
 CC are used to induce a CTL response in a patient. They are also  
 CC useful in compositions for in vivo and ex vivo therapeutic and  
 CC diagnostic applications, e.g. hepatitis B and C.  
 CC infections, e.g. hepatitis B and C.  
 SQ Sequence 9 AA;

Query Match 76.0%; Score 57; DB 1; Length 9;  
 Best Local Similarity 88.9%; Pred. No. 4.40e+01;  
 Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 1 FAMPNFQTL 9  
 ||||| ||  
 QY 1 FAMPNFYTL 9

RESULT 3  
 ID R39631 standard; Protein: 1103 AA.  
 AC R39631.  
 DT 16-DEC-1993 (first entry)  
 DE Neurofibromatosis type 1 polypeptide.  
 KW Non-defective gene; NF-1; treatment; tumours; human; detection; ss.  
 OS Homo sapiens.  
 PN US227292-A.  
 PD 13-JUL-1993.  
 PF 12-JUL-1990; 551531.  
 PR 12-JUL-1990; US-551531.  
 PA (UTAH) UNIV UTAH.  
 PI Cawthon RM, O'Connell P, Viskochil DH, White RL;  
 DR WPI; 93-235118/29.  
 DR N-PSDB; Q46263.  
 PT cDNA encoding neurofibromatosis type 1 gene - for detecting  
 PT defective NF1 genes and tumours caused by such genes  
 PS Disclosure; Fig 1; 59pp; English.  
 CC The sequence is that of the neurofibromatosis type 1 (NF1)  
 CC polypeptide which may be used therapeutically in the treatment of  
 CC diseases associated with defective NF1 genes, e.g. tumours.  
 SQ Sequence 1103 AA;

Query Match 72.0%; Score 54; DB 1; Length 1103;  
 Best Local Similarity 55.6%; Pred. No. 8.69e+01;  
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 551 FSLPKFYLL 559  
 |::|::| |  
 QY 1 FAMPNFYTL 9

RESULT 4  
 ID R59921 standard; protein: 2485 AA.

AC R59921;  
 DT 22-FEB-1995 (first entry)  
 DE RAS associated GAP NF201.  
 KW Ras; GTPase activating protein; GAP; GAP related domain; GRD;  
 KW pK10; pK11; Saccharomyces cerevisiae; RAS2; v-Ras; heat shock;  
 KW neurofibromatosis type 1; NF1.  
 OS Homo sapiens.  
 PN WO9416069-A.  
 PD 21-JUL-1994.  
 PF 12-JAN-1994; U00198.  
 PR 15-JAN-1993; US-004824.  
 PA (SCHE) SCHERING CORP.  
 PI Kaziro Y, Nakafuku M;  
 DR WPI; 94-249216/30.  
 PT Blocking Ras-induced effects on a cell - by introducing a GTPase  
 PT activating protein to the cell, used esp. in treatment of cancers  
 PS Disclosure; Page 36-44; 87pp; English.  
 CC Human neurofibromatosis type 1 (NF1)-GAP related domain (GRD) was  
 CC cloned into the yeast expression vector pK10 to obtain pKp11. The  
 CC pKp11 DNA was mutagenized by hydroxylamine in vitro and transformed  
 CC into S. cerevisiae TK161-R2V-D, which carries an oncogenic-type  
 CC RAS2Val19 mutation. The heat shock sensitivity of the clones was  
 CC checked. Plasmid DNAs were recovered, re-transformed into TK161-  
 CC R2V-D, and phenotypic reversion was examined. 2 Clones, NF201 and  
 CC NF204 (given in R59922), which had strong suppression activity for  
 CC RAS2Val19, were selected. The mutant NF1-GRDs were also able to  
 CC inhibit v-Ras-induced transformation in mammalian cells.  
 SQ Sequence 2485 AA;

Query Match 72.0%; Score 54; DB 1; Length 2485;  
 Best Local Similarity 55.6%; Pred. No. 8.69e+01;  
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 1802 FSLPKFYLL 1810  
 |::|::| |  
 QY 1 FAMPNFYTL 9

RESULT 5  
 ID R59922 standard; protein: 2485 AA.  
 AC R59922;  
 DT 22-FEB-1995 (first entry)  
 DE RAS associated GAP NF204.  
 KW Ras; GTPase activating protein; GAP; GAP related domain; GRD;  
 KW pK10; pK11; Saccharomyces cerevisiae; RAS2; v-Ras; heat shock;  
 KW neurofibromatosis type 1; NF1.  
 OS Homo sapiens.  
 PN WO9416069-A.  
 PD 21-JUL-1994.  
 PF 12-JAN-1994; U00198.  
 PR 15-JAN-1993; US-004824.  
 PA (SCHE) SCHERING CORP.  
 PI Kaziro Y, Nakafuku M;  
 DR WPI; 94-249216/30.  
 PT Blocking Ras-induced effects on a cell - by introducing a GTPase  
 PT activating protein to the cell, used esp. in treatment of cancers  
 PS Disclosure; Page 44-52; 87pp; English.  
 CC Human neurofibromatosis type 1 (NF1)-GAP related domain (GRD) was  
 CC cloned into the yeast expression vector pK10 to obtain pKp11. The  
 CC pKp11 DNA was mutagenized by hydroxylamine in vitro and transformed  
 CC into S. cerevisiae TK161-R2V-D, which carries an oncogenic-type  
 CC RAS2Val19 mutation. The heat shock sensitivity of the clones was  
 CC checked. Plasmid DNAs were recovered, re-transformed into TK161-  
 CC R2V-D, and phenotypic reversion was examined. 2 Clones, NF201  
 CC (given in R59921) and NF204, which had strong suppression activity  
 CC for RAS2Val19, were selected. The mutant NF1-GRDs were also able  
 CC to inhibit v-Ras-induced transformation in mammalian cells.  
 SQ Sequence 2485 AA;

Query Match 72.0%; Score 54; DB 1; Length 2485;  
 Best Local Similarity 55.6%; Pred. No. 8.69e+01;  
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

\*\*\*\*\*  
M P S R E L  
\*\*\*\*\* (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Fri Apr 14 23:30:05 2000; MasPar time 5.26 Seconds  
Tabular output not generated. 40.540 Million cell updates/sec

Title: >US-08-452-843-6  
Description: (1-9) from US08452843.pap  
Perfect Score: 75  
Sequence: 1 FAMPNFYTL 9

Scoring table: PAM 150  
Gap 15

Searched: 188963 seqs, 23686106 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: a-geneseqp  
1:geneseqp

Statistics: Mean 17.224; Variance 58.168; scale 0.296

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	75	100.0	9	1 R89367	Cw3 consensus peptide	6.15e-01
2	57	76.0	9	1 R89366	Cw3 consensus peptide	4.40e-01
3	54	72.0	1103	1 R39631	Neurofibromatosis type	8.69e-01
4	54	72.0	2485	1 R59921	RAS associated GAP NF2	8.69e-01
5	54	72.0	2485	1 R59922	RAS associated GAP NF2	8.69e-01
6	54	72.0	2818	1 W3280	Human neurofibromin.	8.69e-01
7	54	72.0	2818	1 R32268	Nfl gene product.	8.69e-01
8	53	70.7	741	1 R89327	Membrane anchor protei	1.09e-02
9	51	68.0	1464	1 R88469	Feline infectious peri	1.70e-02
10	50	66.7	245	1 W36297	Babesia microti BMNI-1	2.12e-02
11	50	66.7	245	1 W36296	Babesia microti BMNI-1	2.12e-02
12	50	66.7	381	1 W38429	H. pylori GHPO 422 pro	2.12e-02
13	50	66.7	381	1 W5637	H. pylori ORF 03ael080	2.12e-02
14	50	66.7	386	1 W5311	H. pylori ORF 06ep1030	2.12e-02
15	50	66.7	864	1 R33021	Human glucagon degra	2.12e-02
16	49	65.3	409	1 R41227	910 SLG protein.	2.64e-02
17	49	65.3	858	1 R33404	S-Locus receptor (seri	2.64e-02
18	49	65.3	858	1 W49080	Brassica sp. S-recepto	2.64e-02
19	49	65.3	1399	1 R38698	S-PRV-055 TGE virus gp	2.64e-02
20	48	64.0	69	1 Y11991	Human 5' EST secreted	3.29e-02
21	48	64.0	543	1 Y07058	Renal cancer associate	3.29e-02
22	48	64.0	3011	1 R35021	Hepatitis GB virus (HG	3.29e-02
23	47	62.7	28	1 R46636	70 kD proteoglycan cor	4.08e-02

24	47	62.7	39	1 R46637	65 kD proteoglycan cor	4.08e-02
25	47	62.7	217	1 W71251	Protein sequence of th	4.08e-02
26	47	62.7	219	1 W56687	Escherichia coli Car p	4.08e-02
27	47	62.7	219	1 R51279	Chloramphenicol-acetyl	4.08e-02
28	47	62.7	219	1 W56696	Chloramphenicol resist	4.08e-02
29	47	62.7	219	1 R51278	Chloramphenicol-acetyl	4.08e-02
30	47	62.7	240	1 R05425	Amino acid sequence fo	4.08e-02
31	47	62.7	241	1 P92070	Sequence of chloramph	4.08e-02
32	47	62.7	249	1 W38358	Apoptosis associated p	4.08e-02
33	47	62.7	250	1 P92068	Fusion protein compris	4.08e-02
34	47	62.7	251	1 R15611	SP-C from PC210SP-C in	4.08e-02
35	47	62.7	251	1 R05419	CAT:SP-C hybrid protei	4.08e-02
36	47	62.7	293	1 R05418	CAT:SP-B hybrid protei	4.08e-02
37	47	62.7	402	1 P81179	Sequence of human endo	4.08e-02
38	47	62.7	472	1 R56447	TMV replicon-encoded p	4.08e-02
39	47	62.7	485	1 W71249	Protein encoded by rep	4.08e-02
40	47	62.7	930	1 W18061	Pasteurella haemolytic	4.08e-02
41	47	62.7	1854	1 W79161	Human calcium channel	4.08e-02
42	46	61.3	204	1 W47290	Tobacco partial cyto	5.06e-02
43	46	61.3	208	1 W5130	Human secreted protein	5.06e-02
44	46	61.3	594	1 W20603	H. pylori secreted or	5.06e-02
45	46	61.3	919	1 W18580	Potato alpha-glucosida	5.06e-02

ALIGNMENTS

RESULT 1  
ID R89367 standard; peptide; 9 AA.  
AC R89367;  
DT 18-SEP-1996 (first entry)  
DE Cw3 consensus peptide derived immunogenic peptide #2.  
KW Immunogenic peptide; supermotif; HLA molecule; CTL response;  
KW therapeutic; diagnostic; cancer; viral infection; hepatitis B;  
KW hepatitis C.  
OS Synthetic.  
PN WO9603140-A1.  
PD 08-FEB-1996.  
PF 21-JUL-1995; U09234.  
PR 21-JUL-1994; US-278634.  
PR 23-NOV-1994; US-344824.  
PR 30-MAY-1995; US-452843.  
PR (CYTE-) CYTEL CORP.  
PA Sette A, Sidney J;  
PI WPI; 96-116784/12.  
DR Compn. comprising immunogenic peptide with supermotif allowing more  
PT than one HLA mol. to bind - used to induce CTL response in patient  
PT and for in vivo and ex vivo therapeutic and diagnostic applications  
PS Claim 2; Page 26; 32pp; English.  
CC The sequences given in R89362-82 are immunogenic peptides which were  
CC use in the composition of the invention. The composition comprises  
CC an immunogenic peptide of 9-10 residues with a supermotif which  
CC allows binding of more than one HLA molecule. It pref. comprises  
CC two conserved residues, a first at the 2nd position from the N-  
CC terminal is pro, and a 2nd at the C-terminal is Met. These peptides  
CC are used to induce a CTL response in a patient. They are also  
CC useful in compositions for in vivo and ex vivo therapeutic and  
CC diagnostic applications, e.g the treatment of cancer and viral  
CC infections, e.g hepatitis B and C.  
SQ Sequence 9 AA;

Query Match 100.0%; Score 75; DB 1; Length 9;  
Best Local Similarity 100.0%; Pred. No. 6.15e-01;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 FAMPNFYTL 9  
QY 1 FAMPNFYTL 9

RESULT 2  
ID R89366 standard; peptide; 9 AA.  
AC R89366;  
DT 18-SEP-1996 (first entry)

THIS PAGE BLANK (USPTO)

SUMMARY #length 774 #molecular-weight 89513 #checksum 3787  
Query Match 70.0%; Score 49; DB 2; Length 774;  
Best Local Similarity 71.4%; Pred. No. 1.52e+01;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 132 YAMPEFQ 138  
QY 1 FAMPNFQ 7  
:|||||

RESULT 15  
ENTRY #S68146 #type complete  
TITLE AMP deaminase (EC 3.5.4.6), erythrocte, splice form 1a -  
human  
ALTERNATE\_NAMES AMP deaminase isoform E  
CONTAINS AMP deaminase splice form 1b  
ORGANISM #Normal\_name Homo sapiens #common\_name man  
DATE 06-Dec-1996 #sequence\_revision 13-Mar-1997 #text\_change  
17-Mar-1999

ACCESSIONS S68146; S68148; A45071; S28149; S27955  
REFERENCE S68146  
#authors Mahnke-Zizelman, D.K.; Eddy, R.; Shows, T.B.; Sabina, R.L.  
#journal Biochim. Biophys. Acta (1996) 1306:75-92  
#title Characterization of the human AMPD3 gene reveals that 5' exon  
usage is subject to transcriptional control by three  
tandem promoters and alternative splicing.

#cross-references MUID:96201708  
#accession S68146  
#molecule\_type DNA  
#residues 1-776 #label MAH  
#cross-references EMBL:U29925  
#note the nucleotide sequence was submitted to the EMBL Data  
Library, June 1995  
#note only a small part of the nucleic acid sequence is shown  
#note only a small part of the translation is shown  
#note splice form 1a  
#accession S68148  
#status nucleic acid sequence not shown; translation not shown  
#molecule\_type DNA  
#residues 10-776 #label MAW  
#cross-references EMBL:U29925  
#note the nucleotide sequence was submitted to the EMBL Data  
Library, June 1995  
#note splice form 1b; Met-10 is the initiator

REFERENCE A45071  
#authors Mahnke-Zizelman, D.K.; Sabina, R.L.  
#journal J. Biol. Chem. (1992) 267:20866-20877  
#title Cloning of human AMP deaminase isoform E cDNAs. Evidence for  
a third AMPD gene exhibiting alternatively spliced  
5'-exons.

#cross-references MUID:93015995  
#accession A45071  
#molecule\_type mRNA  
#residues 1-216 #label MA2  
#cross-references EMBL:M84720; NID:g178548; PID:g178549  
#note splice form 1a (fragment)  
#note sequence extracted from NCBI backbone (NCBIP:116076)  
#accession B45071  
#status not compared with conceptual translation  
#molecule\_type mRNA  
#residues 10-776 #label MA3  
#cross-references GB:M84721; NID:g178550; PID:g178551  
#note sequence extracted from NCBI backbone (NCBIP:116085)  
#note splice form 1b; Met-10 is the initiator

REFERENCE S28149  
#authors Yamada, Y.; Goto, H.; Ogasawara, N.  
#journal Biochim. Biophys. Acta (1992) 1171:125-128  
#title Cloning and nucleotide sequence of the cDNA encoding human  
erythrocyte-specific AMP deaminase.

#cross-references MUID:93042002  
#accession S28149  
#molecule\_type mRNA

#residues 10-776 #label YAM  
#cross-references GB:D12775; NID:g219456; PID:d1002735; PID:g219457  
#note splice-form 1b; Met-10 is the initiator

GENETICS  
#gene GDB:AMPD3  
#cross-references GDB:136013; OMIM:102772  
#map\_position lp15-1lp15  
#introns 8/1; 83/2; 151/3; 206/1; 279/2; 322/3; 387/3; 431/3; 486/2;  
528/3; 583/2; 623/3; 681/3; 718/3  
CLASSIFICATION #superfamily AMP deaminase  
KEYWORDS alternative initiators; alternative splicing; erythrocyte;  
hydrolase

FEATURE  
1-776 #product AMP deaminase splice form 1a #status predicted  
#label LSPL  
10-776 #product AMP deaminase splice form 1b #status predicted  
#label SSPL  
SUMMARY #length 776 #molecular-weight 89727 #checksum 7740  
Query Match 70.0%; Score 49; DB 2; Length 776;  
Best Local Similarity 71.4%; Pred. No. 1.52e+01;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 134 YAMPEFQ 140  
QY 1 FAMPNFQ 7  
:|||||

Search completed: Fri Apr 14 23:25:11 2000  
Job time : 11 secs.

ENTRY JDLV7 #type complete  
TITLE DNA-directed DNA polymerase (EC 2.7.7.7) - woodchuck  
ORGANISM hepatitis virus (clone 7)  
#formal\_name woodchuck hepatitis virus  
DATE 30-Jun-1989 #sequence\_revision 30-Jun-1989 #text\_change  
12-Jun-1998  
ACCESSIONS C29969  
REFERENCE A94368  
#authors Cohen, J.I.; Miller, R.H.; Rosenblum, B.; Denniston, K.;  
Gerin, J.L.; Purcell, R.H.  
#journal Virology (1988) 162:12-20  
#title Sequence comparison of woodchuck hepatitis virus replicative  
forms shows conservation of the genome.  
#cross-references MUID:88101359  
#accession C29969  
#molecule\_type DNA  
#residues 1-884 #label COH  
#cross-references GB:M18752; NID:g336136; PID:g336138  
CLASSIFICATION #superfamily hepatitis virus DNA-directed DNA polymerase  
KEYWORDS DNA biosynthesis; nucleotidyltransferase  
SUMMARY #length 884 #molecular-weight 99732 #checksum 4231  
Query Match 72.9%; Score 51; DB 1; Length 884;  
Best Local Similarity 77.8%; Pred. NO. 6.24e+00; Indels 0; Gaps 0;  
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
Db 448 FAVPNLQTL 456  
II:II:II  
QY 1 FAMPNFQTL 9  
RESULT 12  
ENTRY JDLVW8 #type complete  
TITLE DNA-directed DNA polymerase (EC 2.7.7.7) - woodchuck  
ORGANISM hepatitis virus (clone 8)  
#formal\_name woodchuck hepatitis virus  
DATE 31-Mar-1990 #sequence\_revision 31-Mar-1990 #text\_change  
12-Jun-1998  
ACCESSIONS A32397  
REFERENCE A94222  
#authors Girones, R.; Cote, P.J.; Hornbuckle, W.E.; Tennant, B.C.;  
Gerin, J.L.; Purcell, R.H.; Miller, R.H.  
#journal Proc. Natl. Acad. Sci. U.S.A. (1989) 86:1846-1849  
#title Complete nucleotide sequence of a molecular clone of  
woodchuck hepatitis virus that is infectious in the natural  
host.  
#cross-references MUID:89184524  
#accession A32397  
#molecule\_type DNA  
#residues 1-884 #label GIR  
#cross-references GB:J04514; NID:g336146  
#note this ORF is not annotated in GenBank entry OHVHEPBA,  
release 106  
GENETICS P  
#gene  
CLASSIFICATION #superfamily hepatitis virus DNA-directed DNA polymerase  
KEYWORDS DNA biosynthesis; nucleotidyltransferase  
SUMMARY #length 884 #molecular-weight 99708 #checksum 2527  
Query Match 72.9%; Score 51; DB 1; Length 884;  
Best Local Similarity 77.8%; Pred. NO. 6.24e+00; Indels 0; Gaps 0;  
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
Db 448 FAVPNLQTL 456  
II:II:II  
QY 1 FAMPNFQTL 9  
RESULT 13  
ENTRY B47050 #type complete  
TITLE glnA 3'-region hypothetical protein - Synecococcus sp.  
ALTERNATE\_NAMES RPD3/accuC homolog  
ORGANISM #formal\_name Synecococcus sp.

DATE 27-Jan-1995 #sequence\_revision 27-Jan-1995 #text\_change  
09-Sep-1997  
ACCESSIONS B47050; S23853  
REFERENCE A47050  
#authors Wagner, S.J.; Thomas, S.P.; Kaufman, R.I.; Nixon, B.T.;  
Stevens Jr., S.E.  
#journal J. Bacteriol. (1993) 175:604-612  
#title The glnA gene of the cyanobacterium Agmenellum quadruplicatum  
PR-6 is nonessential for ammonium assimilation.  
#cross-references MUID:93139025  
#accession B47050  
#status translation not shown  
#molecule\_type DNA  
#residues 1-310 #label WAG  
#cross-references EMBL:213965; NID:g38960; PID:g580726  
#experimental\_source PR-6  
CLASSIFICATION #superfamily RPD3/accuC homology  
FEATURE #domain RPD3/accuC homology #label RAH1  
15-296  
SUMMARY #length 310 #molecular-weight 34145 #checksum 4258  
Query Match 70.0%; Score 49; DB 2; Length 310;  
Best Local Similarity 55.6%; Pred. NO. 1.52e+01;  
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;  
Db 31 FPMKFRLL 39  
II:II:II  
QY 1 FAMPNFQTL 9  
RESULT 14  
ENTRY S68147 #type complete  
TITLE AMP deaminase (EC 3.5.4.6), erythrocte, splice form 1c -  
human  
ALTERNATE\_NAMES AMP deaminase isoform E  
ORGANISM #formal\_name Homo sapiens #common\_name man  
DATE 06-Dec-1996 #sequence\_revision 13-Mar-1997 #text\_change  
17-Mar-1999  
ACCESSIONS S68147; C45071  
REFERENCE S68146  
#authors Mahake-Zizelman, D.K.; Eddy, R.; Shows, T.B.; Sabina, R.L.  
#journal Biochim. Biophys. Acta (1996) 1306:75-92  
#title Characterization of the human AMPD3 gene reveals that 5' exon  
usage is subject to transcriptional control by three  
tandem promoters and alternative splicing.  
#cross-references MUID:96201708  
#accession S68147  
#molecule\_type DNA  
#residues 1-774 #label MAH  
#cross-references EMBL:U29925  
#note the nucleotide sequence was submitted to the EMBL Data  
Library, June 1995  
#note only a small part of the translation is shown  
#note only a small part of the translation is shown  
REFERENCE A45071  
#authors Mahake-Zizelman, D.K.; Sabina, R.L.  
#journal J. Biol. Chem. (1992) 267:20866-20877  
#title Cloning of human AMP deaminase isoform E cDNAs. Evidence for  
a third AMPD gene exhibiting alternatively spliced  
5'-exons  
#cross-references MUID:93015995  
#accession C45071  
#status preliminary; not compared with conceptual translation  
#molecule\_type nucleic acid  
#residues 1-658 #label MA2  
#cross-references GB:M84722; NID:G178552; PID:g553179  
#note sequence extracted from NCBI backbone (NCBIP:116090)  
GENETICS  
#gene GDB:AMPD3  
#introns 6/1; 81/2; 149/3; 204/1; 277/2; 320/3; 385/3; 429/3; 484/2;  
CLASSIFICATION 526/3; 581/2; 621/3; 679/3; 716/3  
#superfamily AMP deaminase  
KEYWORDS alternative splicing; erythrocyte; hydrolase

##residues 1-625 #label KAW  
##cross-references GB:AP000003; NID:g3236130; PID:d1030733; PID:g3257107  
##experimental\_source strain OT3  
##note this accession replaces an interim accession for a  
sequence replaced by GenBank

## GENETICS

#gene PH0699  
CLASSIFICATION #superfamily Methanococcus jannaschii threonine--tRNA ligase  
KEYWORDS aminoacyl-tRNA synthetase; protein biosynthesis  
SUMMARY #length 625 #molecular-weight 73023 #checksum 7363

Query Match 72.9%; Score 51; DB 2; Length 625;  
Best Local Similarity 55.6%; Pred. No. 6.24e+00;  
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 344 FTMPDMHTL 352

QY 1 FAMPNFQTL 9

RESULT 7 JDVLC #type complete  
ENTRY DNA-directed DNA polymerase (EC 2.7.7.7) - woodchuck  
TITLE hepatitis virus (clone 1)  
ORGANISM #formal\_name woodchuck hepatitis virus  
DATE 14-Nov-1983 #sequence\_revision 14-Nov-1983 #text\_change  
12-Jun-1998

ACCESSIONS A00707  
REFERENCE A92986  
#authors Galibert, F.; Chen, T.N.; Mandart, E.  
#journal J. Virol. (1982) 41:51-65  
#title Nucleotide sequence of a cloned woodchuck hepatitis virus  
genome: comparison with the hepatitis B virus sequence.  
#cross-references MUID:82216969  
#accession A00707

##molecule\_type DNA  
##residues 1-879 #label GAL  
##cross-references GB:J02442; NID:g336126; PID:g336127  
CLASSIFICATION #superfamily hepatitis virus DNA-directed DNA polymerase  
KEYWORDS DNA biosynthesis; nucleotidyltransferase  
SUMMARY #length 879 #molecular-weight 99185 #checksum 8623

Query Match 72.9%; Score 51; DB 1; Length 879;  
Best Local Similarity 77.8%; Pred. No. 6.24e+00;  
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 443 FAVPNLQTL 451

QY 1 FAMPNFQTL 9

RESULT 8 JDVLS #type complete  
ENTRY DNA-directed DNA polymerase (EC 2.7.7.7) - ground squirrel  
TITLE hepatitis virus  
ORGANISM #formal\_name ground squirrel hepatitis virus  
DATE 25-Feb-1985 #sequence\_revision 25-Feb-1985 #text\_change  
20-Mar-1998

ACCESSIONS A00709  
REFERENCE A93000  
#authors Seeger, C.; Ganem, D.; Varmus, H.E.  
#journal J. Virol. (1984) 51:367-375  
#title Nucleotide sequence of an infectious molecularly cloned  
genome of ground squirrel hepatitis virus.  
#cross-references MUID:84267998  
#accession A00709

##molecule\_type DNA  
##residues 1-881 #label SEE  
##cross-references GB:K02715; NID:g325400; PID:g325402  
CLASSIFICATION #superfamily hepatitis virus DNA-directed DNA polymerase  
KEYWORDS DNA biosynthesis; nucleotidyltransferase  
SUMMARY #length 881 #molecular-weight 99976 #checksum 6194

Query Match 72.9%; Score 51; DB 1; Length 881;  
Best Local Similarity 77.8%; Pred. No. 6.24e+00;  
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 445 FAVPNLQTL 453

QY 1 FAMPNFQTL 9

RESULT 9 JDVLC2 #type complete  
ENTRY DNA-directed DNA polymerase (EC 2.7.7.7) - woodchuck  
TITLE hepatitis virus (clone 2)  
ORGANISM #formal\_name woodchuck hepatitis virus  
DATE 30-Jun-1987 #sequence\_revision 30-Jun-1987 #text\_change  
12-Jun-1998

ACCESSIONS A00708  
REFERENCE A93015  
#authors Kodama, K.; Ogasawara, N.; Yoshikawa, H.; Murakami, S.  
#journal J. Virol. (1985) 56:978-986  
#title Nucleotide sequence of a cloned woodchuck hepatitis virus  
genome: evolutionary relationship between hepadnaviruses.  
#cross-references MUID:86062931  
#accession A00708

##molecule\_type DNA  
##residues 1-883 #label KOD  
##cross-references GB:M1082; NID:g336132; PID:g336134  
CLASSIFICATION #superfamily hepatitis virus DNA-directed DNA polymerase  
KEYWORDS DNA biosynthesis; nucleotidyltransferase  
SUMMARY #length 883 #molecular-weight 99346 #checksum 593

Query Match 72.9%; Score 51; DB 1; Length 883;  
Best Local Similarity 77.8%; Pred. No. 6.24e+00;  
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 447 FAVPNLQTL 455

QY 1 FAMPNFQTL 9

RESULT 10 JDVLS9 #type complete  
ENTRY DNA-directed DNA polymerase (EC 2.7.7.7) - woodchuck  
TITLE hepatitis virus (clone 59)  
ORGANISM #formal\_name woodchuck hepatitis virus  
DATE 30-Jun-1989 #sequence\_revision 30-Jun-1989 #text\_change  
12-Jun-1998

ACCESSIONS G29969  
REFERENCE A94368  
#authors Cohen, J.I.; Miller, R.H.; Rosenblum, B.; Denniston, K.;  
Gerin, J.L.; Purcell, R.H.  
#journal Virology (1988) 162:12-20  
#title Sequence comparison of woodchuck hepatitis virus replicative  
forms shows conservation of the genome.  
#cross-references MUID:88101359  
#accession G29969

##molecule\_type DNA  
##residues 1-884 #label COH  
##cross-references GB:M19183; NID:g336141; PID:g336143  
CLASSIFICATION #superfamily hepatitis virus DNA-directed DNA polymerase  
KEYWORDS DNA biosynthesis; nucleotidyltransferase  
SUMMARY #length 884 #molecular-weight 99399 #checksum 3128

Query Match 72.9%; Score 51; DB 1; Length 884;  
Best Local Similarity 77.8%; Pred. No. 6.24e+00;  
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 448 FAVPNLQTL 456

QY 1 FAMPNFQTL 9

RESULT 11



```

GENETICS
#map_position FOR418436-419089
CLASSIFICATION #superfamily Escherichia coli ribosomal protein S5
KEYWORDS #length 744 #molecular-weight 80266 #checksum 9998
SUMMARY

Query Match 75.7%; Score 53; DB 2; Length 744;
Best Local Similarity 55.6%; Pred. No. 2.49e+00;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 519 FEMPDFFTF 527
|||:|:|:
QY 1 FAMPNFQTL 9

RESULT 3
ENTRY S72830 #type complete
TITLE hypothetical protein B1620_F1.14 - Mycobacterium leprae
ORGANISM #formal_name Mycobacterium leprae
DATE 19-Mar-1997 #sequence_revision 25-Apr-1997 #text_change
09-Sep-1997
ACCESSIONS S72830
REFERENCE Smith, D.R.; Robison, K.
#authors
#submission Submitted to the EMBL Data Library, November 1993
#description Mycobacterium leprae cosmid B1620.
#accession S72830
#status preliminary
#molecule_type DNA
#residues 1-87 #label SMI
#cross-references EMBL:U00015; NID:g466931; PID:g466951
GENETICS
#start_codon GTG
SUMMARY #length 87 #molecular-weight 9272 #checksum 188

Query Match 74.3%; Score 52; DB 2; Length 87;
Best Local Similarity 55.6%; Pred. No. 3.96e+00;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 19 FGMTNFQAM 27
|||:|:|:
QY 1 FAMPNFQTL 9

RESULT 4
ENTRY C64359 #type complete
TITLE ribosomal protein S5 - Methanococcus jannaschii
ORGANISM #formal_name Methanococcus jannaschii
DATE 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change
10-Oct-1997
ACCESSIONS C64359
REFERENCE A64300
#authors Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake, J.A.; FitzGerald, L.M.; Clayton, R.A.; Gocayne, J.D.; Kerlavage, A.R.; Dougherty, B.A.; Tomb, J.F.; Adams, M.D.; Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.; Scott, J.L.; Geoghegan, N.S.M.; Weidman, J.F.; Fuhrmann, J.L.; Nguyen, D.; Uitterback, T.R.; Kelley, J.M.; Peterson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.; Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese, C.R.; Venter, J.C. Science (1996) 273:1058-1073
#journal Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii.
#title
#cross-references MUID:96337999
#accession C64359
#status preliminary; nucleic acid sequence not shown; translation not shown
#molecule_type DNA
#residues 1-217 #label BUL
#cross-references GB:U67497; GB:L77117; NID:g1591160; PID:g1591177; TIGR:MJ0475; PID:g1510548

residue 1 as Val

GENETICS
#map_position FOR418436-419089
CLASSIFICATION #superfamily Escherichia coli ribosomal protein S5
KEYWORDS #length 744 #molecular-weight 80266 #checksum 9998
SUMMARY

Query Match 75.7%; Score 53; DB 2; Length 744;
Best Local Similarity 55.6%; Pred. No. 2.49e+00;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 519 FEMPDFFTF 527
|||:|:|:
QY 1 FAMPNFQTL 9

RESULT 3
ENTRY S72830 #type complete
TITLE hypothetical protein B1620_F1.14 - Mycobacterium leprae
ORGANISM #formal_name Mycobacterium leprae
DATE 19-Mar-1997 #sequence_revision 25-Apr-1997 #text_change
09-Sep-1997
ACCESSIONS S72830
REFERENCE Smith, D.R.; Robison, K.
#authors
#submission Submitted to the EMBL Data Library, November 1993
#description Mycobacterium leprae cosmid B1620.
#accession S72830
#status preliminary
#molecule_type DNA
#residues 1-87 #label SMI
#cross-references EMBL:U00015; NID:g466931; PID:g466951
GENETICS
#start_codon GTG
SUMMARY #length 87 #molecular-weight 9272 #checksum 188

Query Match 74.3%; Score 52; DB 2; Length 87;
Best Local Similarity 55.6%; Pred. No. 3.96e+00;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 19 FGMTNFQAM 27
|||:|:|:
QY 1 FAMPNFQTL 9

RESULT 4
ENTRY C64359 #type complete
TITLE ribosomal protein S5 - Methanococcus jannaschii
ORGANISM #formal_name Methanococcus jannaschii
DATE 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change
10-Oct-1997
ACCESSIONS C64359
REFERENCE A64300
#authors Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake, J.A.; FitzGerald, L.M.; Clayton, R.A.; Gocayne, J.D.; Kerlavage, A.R.; Dougherty, B.A.; Tomb, J.F.; Adams, M.D.; Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.; Scott, J.L.; Geoghegan, N.S.M.; Weidman, J.F.; Fuhrmann, J.L.; Nguyen, D.; Uitterback, T.R.; Kelley, J.M.; Peterson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.; Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese, C.R.; Venter, J.C. Science (1996) 273:1058-1073
#journal Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii.
#title
#cross-references MUID:96337999
#accession C64359
#status preliminary; nucleic acid sequence not shown; translation not shown
#molecule_type DNA
#residues 1-217 #label BUL
#cross-references GB:U67497; GB:L77117; NID:g1591160; PID:g1591177; TIGR:MJ0475; PID:g1510548

GENETICS
#map_position FOR418436-419089
CLASSIFICATION #superfamily Escherichia coli ribosomal protein S5
KEYWORDS #length 217 #molecular-weight 23839 #checksum 5006
SUMMARY

Query Match 72.9%; Score 51; DB 2; Length 217;
Best Local Similarity 55.8%; Pred. No. 6.24e+00;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 185 FAMATFEAL 193
|||:|:|:
QY 1 FAMPNFQTL 9

RESULT 5
ENTRY JDVL64 #type fragment
TITLE DNA-directed DNA polymerase (EC 2.7.7.7) - woodchuck hepatitis virus (clone 64) (fragment)
ORGANISM #formal_name woodchuck hepatitis virus
DATE 31-Mar-1989 #sequence_revision 31-Mar-1989 #text_change
12-Jun-1998
ACCESSIONS A29498
REFERENCE Etienne, J.; Moeroy, T.; Trepo, C.; Tiollais, P.; Buendia, M.A.
#authors
#journal Gene (1986) 50:207-214
#title Nucleotide sequence of the woodchuck hepatitis virus surface antigen mRNAs and the variability of three overlapping viral genes.
#cross-references MUID:87219879
#accession A29498
#molecule_type mRNA
#residues 1-556 #label ETI
#cross-references GB:M15954; NID:g893289; PID:g336155
GENETICS
#gene P
CLASSIFICATION #superfamily hepatitis virus DNA-directed DNA polymerase
KEYWORDS DNA biosynthesis; nucleotidyltransferase
SUMMARY #length 556 #checksum 3238

Query Match 72.9%; Score 51; DB 1; Length 556;
Best Local Similarity 77.8%; Pred. No. 6.24e+00;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 120 FAVPNEQTL 128
|||:|:|:
QY 1 FAMPNFQTL 9

RESULT 6
ENTRY D71116 #type complete
TITLE Probable threonyl-tRNA synthetase - Pyrococcus horikoshii
ORGANISM #formal_name Pyrococcus horikoshii
DATE 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change
21-Nov-1998
ACCESSIONS D71116
REFERENCE A71000
#authors Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Sekine, M.; Baba, S.; Kosugi, H.; Hosoyama, A.; Nagai, Y.; Sakai, M.; Ogura, K.; Otsuka, R.; Nakazawa, H.; Takamiya, M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi, A.; Aoki, K.; Yoshizawa, T.; Nakamura, Y.; Robb, F.T.; Horikoshi, K.; Masuchi, Y.; Shizuya, H.; Kikuchi, H. DNA Res. (1998) 5:55-76
#journal Complete sequence and gene organization of the genome of a hyper-thermophilic archaeobacterium, Pyrococcus horikoshii OT3.
#title
#cross-references MUID:98344137
#accession D71116
#status preliminary; nucleic acid sequence not shown; translation not shown
#molecule_type DNA
```

\*\*\*\*\*  
M I S E R E  
\*\*\*\*\* (TM)  
\*\*\*\*\*  
Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd  
MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Fri Apr 14 23:25:00 2000; MasPar time 3.33 Seconds  
Tabular output not generated.  
108.286 Million cell updates/sec

Title: >US-08-452-843-5  
Description: (1-9) from US08452843.pep  
Perfect Score: 70  
Sequence: 1 FAMPNFQTL 9  
Scoring table: PAM 150  
Gap 15  
Searched: 122810 seqs, 40068593 residues  
Post-processing: Minimum Match 0%  
Listing first 45 summaries  
Database: p1r2  
1:p1r1 2:p1r2 3:p1r3 4:p1r4  
Statistics: Mean 23.396; Variance 30.146; scale 0.776  
Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.  
SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	62	88.6	952	2	T03158 tegument protein 63 -	3.11e-02
2	53	75.7	744	2	S65669 biotin sulfoxide redu	2.49e+00
3	52	74.3	87	2	S72830 hypothetical protein	3.96e+00
4	51	72.9	217	2	C64359 ribosomal protein S5	6.24e+00
5	51	72.9	556	1	JDVL64 DNA-directed DNA poly	6.24e+00
6	51	72.9	625	2	D71116 probable threonyl-trn	6.24e+00
7	51	72.9	879	1	JDVLC DNA-directed DNA poly	6.24e+00
8	51	72.9	881	1	JDVLS DNA-directed DNA poly	6.24e+00
9	51	72.9	883	1	JDVLC2 DNA-directed DNA poly	6.24e+00
10	51	72.9	884	1	JDVL59 DNA-directed DNA poly	6.24e+00
11	51	72.9	884	1	JDVL7 DNA-directed DNA poly	6.24e+00
12	51	72.9	884	1	JDVLW8 DNA-directed DNA poly	6.24e+00
13	49	70.0	310	2	B47050 glnA 3'-region hypoth	1.52e+01
14	49	70.0	774	2	S68147 AMP deaminase (EC 3.5	1.52e+01
15	49	70.0	776	2	S68146 AMP deaminase (EC 3.5	1.52e+01
16	48	68.6	205	2	I49365 protein tyrosine phos	2.36e+01
17	48	68.6	223	2	I49365 protein tyrosine phos	2.36e+01
18	48	68.6	564	1	HMIVF3 hemagglutinin precurs	2.36e+01
19	48	68.6	564	1	HMIVF6 hemagglutinin precurs	2.36e+01
20	48	68.6	564	1	HMIVF9 hemagglutinin precurs	2.36e+01
21	48	68.6	564	1	HMIVF5 hemagglutinin precurs	2.36e+01
22	48	68.6	564	1	HMIVF4 hemagglutinin precurs	2.36e+01
23	48	68.6	564	1	HMIVF8 hemagglutinin precurs	2.36e+01

24 48 68.6 568 1 A46339 hemagglutinin precurs 2.36e+01  
25 48 68.6 567 1 BHLOA hemocyanin chain a - 2.36e+01  
26 48 68.6 567 1 BHLOV hemocyanin chain b - 2.36e+01  
27 48 68.6 750 1 JDVLVH DNA-directed DNA poly 2.36e+01  
28 48 68.6 832 1 JDVLVA DNA-directed DNA poly 2.36e+01  
29 48 68.6 832 1 S20752 DNA-directed DNA poly 2.36e+01  
30 48 68.6 832 2 S71785 DNA-directed DNA poly 2.36e+01  
31 48 68.6 832 1 JDVLC2 DNA-directed DNA poly 2.36e+01  
32 48 68.6 832 1 JDVLA1 DNA-directed DNA poly 2.36e+01  
33 48 68.6 832 2 S67505 DNA-directed DNA poly 2.36e+01  
34 48 68.6 832 1 S47405 DNA-directed DNA poly 2.36e+01  
35 48 68.6 832 1 JDVLVB DNA-directed DNA poly 2.36e+01  
36 48 68.6 832 1 S20757 DNA-directed DNA poly 2.36e+01  
37 48 68.6 842 1 JDVLVS DNA-directed DNA poly 2.36e+01  
38 48 68.6 843 1 JDVLJ3 DNA-directed DNA poly 2.36e+01  
39 48 68.6 843 1 JQ2229 DNA-directed DNA poly 2.36e+01  
40 48 68.6 843 1 JDVLJ1 DNA-directed DNA poly 2.36e+01  
41 48 68.6 843 1 S43491 DNA-directed DNA poly 2.36e+01  
42 48 68.6 843 1 JDVLJ2 DNA-directed DNA poly 2.36e+01  
43 48 68.6 845 1 JDVLKS DNA-directed DNA poly 2.36e+01  
44 48 68.6 845 1 JDVLVD DNA-directed DNA poly 2.36e+01  
45 48 68.6 967 2 C70831 probable mmpL4 protei 2.36e+01

ALIGNMENTS

RESULT 1  
ENTRY T03158 #type complete  
TITLE tegument protein 63 - aicelaphine herpesvirus 1  
ORGANISM #formal\_name aicelaphine herpesvirus 1  
DATE 24-Mar-1999 #sequence\_revision 24-Mar-1999 #text\_change 24-Mar-1999  
ACCESSIONS T03158  
REFERENCE Z14840  
#authors Ensser, A.; Pflanz, R.; Fleckenstein, B.  
#journal J. Virol. (1997) 71:6517-6525  
#title Primary structure of the aicelaphine herpesvirus 1 genome.  
#accession T03158  
#status Preliminary; translated from GB/EMBL/DBDJ  
#molecule\_type DNA  
#residues 1-952 #label ENS  
#cross-references EMBL:AF005370; NID:g2337967; PID:g2338026  
SUMMARY #length 952 #molecular-weight 107065 #checksum 7079  
Query Match 88.6%; Score 62; DB 2; Length 952;  
Best Local Similarity 77.8%; Pred No. 3.11e-02;  
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 185 FMPNFQTM 193  
QY 1 FAMPNFQTL 9  
RESULT 2  
ENTRY S65669 #type complete  
TITLE biotin sulfoxide reductase (EC 1.8.4.-) - Rhodobacter  
ORGANISM sphaeroides  
DATE #formal\_name Rhodobacter sphaeroides  
14-Feb-1997 #sequence\_revision 13-Mar-1997 #text\_change 24-Oct-1998  
ACCESSIONS S65669  
REFERENCE S65669  
#authors Pollock, V.V.; Barber, M.J.  
#journal Arch. Biochem. Biophys. (1995) 318:322-332  
#title Molecular cloning and expression of biotin sulfoxide  
reductase from Rhodobacter sphaeroides forma Sp.  
denitrificans.  
#accession S65669  
#molecule\_type DNA  
#residues 1-744 #label POL  
#cross-references EMBL:U08189; NID:g953223; PID:g953224  
#experimental\_source strain forma sp. denitrificans  
#note the authors translated the initiation codon GTG for

THIS PAGE BLANK (USPTO)

QY 1 YPAEITLYW 9  
 !!!!!!!

RESULT 15  
 ID Q30444 PRELIMINARY; PRT; 356 AA.  
 AC Q30444;  
 DT 01-NOV-1996 (Tremblrel. 01, Created)  
 DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)  
 DT 01-NOV-1999 (Tremblrel. 12, Last annotation update)  
 DE MHC CLASS I CAJA-G\*04 (FRAGMENT).  
 GN CAJA-G.  
 OS Callithrix jacchus (Common marmoset).  
 OC Eukaryota; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Primates; Platyrrhini; Callitrichidae; Callithrix.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 98070787.  
 RA CADAVID L.F.; SHUFFLEBOTHAM C.; RUIZ F.J.; YEAGER M.; HUGHES A.L.;  
 RA WATKINS D.I.;  
 RT "Evolutionary instability of the major histocompatibility complex  
 class I loci in New World primates."  
 RL Proc. Natl. Acad. Sci. U.S.A. 94:14536-14541(1997).  
 DR EMBL; U59640; AAB97483.1; -.  
 DR HSSP; P30491; IAIM.  
 DR PFAM; PF00047; IG; 1.  
 DR PFAM; PF00129; MHC\_I; 1.  
 KW MHC.  
 FT NON\_TER 1  
 SQ SEQUENCE 356 AA; 39894 MW; ECF4FAE CRC32;

Query Match 80.0%; Score 64; DB 7; Length 356;  
 Best Local Similarity 88.9%; Pred. No. 5.85e-01;  
 Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 225 YPAEITLYW 233  
 !!!!!!!  
 QY 1 YPAEITLYW 9

Search completed: Fri Apr 14 23:22:17 2000  
 Job time : 102 secs.



```
DR PROSITE; PS00290; IG_MHC; 1.
DR PFAM; PF00047; Ig; 1.
DR PFAM; PF00129; MHC_I; 1.
KW MHC.
FT NON_TER 298 298 34428 MW; E1126994 CRC32;
SQ SEQUENCE 298 AA; 34428 MW; 61126994 CRC32;

Query Match
Best Local Similarity 80.0%; Score 64; DB 7; Length 298;
Best Local Similarity 88.9%; Pred. No. 5.85e-01;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 233 YPAEITLW 241
   |||||
QY 1 YPAEITLW 9

RESULT 7
ID Q95351 PRELIMINARY; PRT; 298 AA.
AC Q95351;
DT 01-FEB-1997 (TrEMBLrel. 02, Created)
DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE HLA-B*1529 (FRAGMENT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-BLOOD;
RX MEDLINE; 96369309.
RA LIN L., TOKUNAGA K., TANAKA H., NAKAJIMA F., IMANISHI T.,
RA KASHIWASE K., BANNAI M., MIZUNO S., AKAZA T., TADOKORO K., SHIBATA Y.,
RA JUJI T.;
RT "Further molecular diversity in the HLA-B*15 group.";
RL Tissue Antigens 47:265-274(1996).
DR EMBL; D44501; BAA07944.1; -.
DR HSSP; P30685; 1A9E.
DR PROSITE; PS00290; IG_MHC; 1.
DR PFAM; PF00047; Ig; 1.
DR PFAM; PF00129; MHC_I; 1.
FT NON_TER 298 298 34300 MW; 6F8C155C CRC32;
SQ SEQUENCE 298 AA; 34300 MW; 6F8C155C CRC32;

Query Match
Best Local Similarity 80.0%; Score 64; DB 7; Length 298;
Best Local Similarity 88.9%; Pred. No. 5.85e-01;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 233 YPAEITLW 241
   |||||
QY 1 YPAEITLW 9

RESULT 8
ID Q29948 PRELIMINARY; PRT; 300 AA.
AC Q29948;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE CELL SURFACE ANTIGEN (FRAGMENT).
GN HLA-A.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 84287690.
RA ARNOT D., LILLIE J.W., AUFFRAY C., KAPPEL D., STROMINGER J.L.;
RT "Inter-locus and intra-allelic polymorphisms of HLA class I antigen
RT gene mRNA.";
RL Immunogenetics 20:237-252(1984).
DR EMBL; M27540; AAA59638.1; -.
DR HSSP; P30460; 1A9E.
DR PROSITE; PS00290; IG_MHC; 1.

DR PROSITE; PS00290; IG_MHC; 1.
DR PFAM; PF00047; Ig; 1.
DR PFAM; PF00129; MHC_I; 1.
KW MHC.
FT NON_TER 300 AA; 33515 MW; 62CD8543 CRC32;
SQ SEQUENCE 300 AA; 33515 MW; 62CD8543 CRC32;

Query Match
Best Local Similarity 80.0%; Score 64; DB 7; Length 300;
Best Local Similarity 88.9%; Pred. No. 5.85e-01;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 171 YPAEITLW 179
   |||||
QY 1 YPAEITLW 9

RESULT 9
ID Q29656 PRELIMINARY; PRT; 322 AA.
AC Q29656;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE HLA-B*71 VARIANT (FRAGMENT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RA HURLEY C.K., BEI M., RODRIGUEZ S., JOHNSON A.;
RL Submitted (JUN-1994) to the EMBL/GenBank/DBJ databases.
DR EMBL; U11268; AA19928.1; -.
DR HSSP; P30685; 1A9E.
DR PROSITE; PS00290; IG_MHC; 1.
DR PFAM; PF00047; Ig; 1.
DR PFAM; PF00129; MHC_I; 1.
KW MHC.
FT NON_TER 322 322 36558 MW; 4707255F CRC32;
SQ SEQUENCE 322 AA; 36558 MW; 4707255F CRC32;

Query Match
Best Local Similarity 80.0%; Score 64; DB 7; Length 322;
Best Local Similarity 88.9%; Pred. No. 5.85e-01;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 233 YPAEITLW 241
   |||||
QY 1 YPAEITLW 9

RESULT 10
ID Q29654 PRELIMINARY; PRT; 322 AA.
AC Q29654;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE HLA-B*71 (FRAGMENT).
GN B-1510.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RA HURLEY C.K., BEI M., RODRIGUEZ S., JOHNSON A.;
RL Submitted (JUN-1994) to the EMBL/GenBank/DBJ databases.
DR EMBL; U11264; AA19924.1; -.
DR HSSP; P30685; 1A9E.
DR PROSITE; PS00290; IG_MHC; 1.
DR PFAM; PF00047; Ig; 1.
DR PFAM; PF00129; MHC_I; 1.
KW MHC.
FT NON_TER 322 322 36634 MW; 7D650C0C CRC32;
SQ SEQUENCE 322 AA; 36634 MW; 7D650C0C CRC32;

Query Match
Best Local Similarity 80.0%; Score 64; DB 7; Length 322;
Best Local Similarity 88.9%; Pred. No. 5.85e-01;
```

```

01-MAY-1997 (TREMBlrel. 03, Created)
01-MAY-1997 (TREMBlrel. 03, Last sequence update)
01-NOV-1999 (TREMBlrel. 12, Last annotation update)
MHC CLASS I PLA-A1 ALPHA-CHAIN (FRAGMENT).
Phoca vitulina (Harbor seal).
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Carnivora; Pinnipedia; Phocidae; Phoca.
[1]
SEQUENCE FROM N.A.
ZHONG J.F., BOOTHBY J.;
Immunogenetics 0:0-0(1998).
EMBL; U88874; AAC83173.1; -.
HSSP; P30685; 1A9E.
DR PROSITE; PS00290; IG_MHC; 1.
DR PFAM; PF00047; Ig; 1.
DR PFAM; PF00129; MHC_I; 1.
KW MHC.
FT NON_TER 1 1
FT NON_TER 335 335
SQ SEQUENCE 335 AA; 37840 MW; 18BA05C3 CRC32;

Query Match 81.3%; Score 65; DB 7; Length 335;
Best Local Similarity 88.9%; Pred. No. 3.84e-01;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 208 YPAEITLW 216
|||||||
QY 1 YPAEITLW 9

RESULT 3
ID Q30896 PRELIMINARY; PRT; 355 AA.
AC Q30896;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
DE MHC CLASS I PIP1-G*03 (FRAGMENT).
GN PIP1-G.
OS Pithecia pithecia (White-faced saki).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Platyrrhini; Cebidae; Pitheciinae; Pithecia.
[1]
SEQUENCE FROM N.A.
MEDLINE; 98070787.
RA CADAVID L.F., SHUFFLEBOTHAM C., RUIZ F.J., YEAGER M., HUGHES A.L.,
WATKINS D.I.;
RT "Evolutionary instability of the major histocompatibility complex
class I loci in New World primates";
RL Proc. Natl. Acad. Sci. U.S.A. 94:14536-14541(1997).
DR EMBL; U59653; AAB97496.1; -.
DR HSSP; P30685; 1A9E.
DR PROSITE; PS00290; IG_MHC; 1.
DR PFAM; PF00047; Ig; 1.
DR PFAM; PF00129; MHC_I; 1.
KW MHC.
FT NON_TER 1 1
FT NON_TER 355 355
SQ SEQUENCE 355 AA; 40137 MW; 0FF8C93 CRC32;

Query Match 81.3%; Score 65; DB 7; Length 355;
Best Local Similarity 88.9%; Pred. No. 3.84e-01;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 225 YPAEITLW 233
|||||||
QY 1 YPAEITLW 9

RESULT 4
ID Q9XRN4 PRELIMINARY; PRT; 157 AA.
AC Q9XRN4;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)

DE MHC CLASS I ANTIGEN (FRAGMENT).
OS Rhinoceros unicornis (Greater Indian rhinoceros).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Perissodactyla; Rhinocerotidae; Rhinoceros.
[1]
SEQUENCE FROM N.A.
TISSUE-PERIPHERAL BLOOD;
RA HOLMES E.C., ELLIS S.A.;
RT "Evolutionary History of MHC class I genes in the mammalian order
Perissodactyla";
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ133684; CAB42821.1; -.
FT NON_TER 1 1
FT NON_TER 157 157
SQ SEQUENCE 157 AA; 17290 MW; 317358F9 CRC32;

Query Match 80.0%; Score 64; DB 7; Length 157;
Best Local Similarity 88.9%; Pred. No. 5.85e-01;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 26 YPAEITLW 34
|||||||
QY 1 YPAEITLW 9

RESULT 5
ID P79542 PRELIMINARY; PRT; 225 AA.
AC P79542;
DT 01-MAY-1997 (TREMBlrel. 03, Created)
DT 01-MAY-1997 (TREMBlrel. 03, Last sequence update)
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
DE MHC CLASS I ANTIGEN (FRAGMENT).
GN HLA-B.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
[1]
SEQUENCE FROM N.A.
EBERLE M., LORENTZEN D., IWANAGA K.K., WATKINS D.I.;
RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; U74386; AAB41720.1; -.
DR HSSP; P30460; 1AGE.
DR PFAM; PF00129; MHC_I; 1.
KW MHC.
FT NON_TER 1 1
FT NON_TER 225 225
SQ SEQUENCE 225 AA; 25906 MW; 136BC712 CRC32;

Query Match 80.0%; Score 64; DB 7; Length 225;
Best Local Similarity 88.9%; Pred. No. 5.85e-01;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 201 YPAEITLW 209
|||||||
QY 1 YPAEITLW 9

RESULT 6
ID O19657 PRELIMINARY; PRT; 298 AA.
AC O19657;
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
DE HLA-CW*0602 (FRAGMENT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
[1]
SEQUENCE FROM N.A.
TISSUE-PERIPHERAL BLOOD;
RA WANG H., TOKUNAGA K.;
RL Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; D64147; BAA19533.1; -.
DR HSSP; P30460; 1AGE.
```

\*\*\*\*\*  
 W P S R E L L  
 \*\*\*\*\*  
 (TM)

Release 3.1A John F. Collins, BioComputing Research Unit.  
 Copyright (c) 1993-1998 University of Edinburgh, U.K.  
 Distribution rights by Oxford Molecular Ltd

MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm  
 Run on: Fri Apr 14 23:20:35 2000; MasPar time 13.46 Seconds  
 Tabular output not generated. 46.345 Million cell updates/sec

Title: >US-08-452-843-4  
 Description: (1-9) from US08452843.pep  
 Perfect Score: 80  
 Sequence: 1 YPAEITLYW 9

Scoring table: PAM 150  
 Gap 15

Searched: 225878 seqs, 69334122 residues

Post-processing: Minimum Match 0%  
 Listing first 45 summaries

Database: sptrembl12  
 1:sp.archaea 2:sp.bacteria 3:sp.fungi 4:sp.human  
 5:sp.invertebrate 6:sp.mammal 7:sp.mhc 8:sp.organelle  
 9:sp.phage 10:sp.plant 11:sp.rodent 12:sp.unclassified  
 13:sp.vertebrate 14:sp.virus

Statistics: Mean 25.150; Variance 38.453; scale 0.654

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	68	85.0	1226	1 Q38836	HYPOTHETICAL PROTEIN M	1.06e-01
2	65	81.3	335	7 P79607	MHC CLASS I PLA-A1 ALP	3.84e-01
3	65	81.3	355	7 Q30896	MHC CLASS I PIP1-G*03	3.84e-01
4	64	80.0	157	7 Q30894	MHC CLASS I ANTIGEN (F)	5.85e-01
5	64	80.0	225	7 P79542	MHC CLASS I ANTIGEN (F)	5.85e-01
6	64	80.0	298	7 Q39657	HLA-CW*0602 (FRAGMENT)	5.85e-01
7	64	80.0	298	7 Q39351	HLA-B*1529 (FRAGMENT)	5.85e-01
8	64	80.0	300	7 Q29948	CELL SURFACE ANTIGEN (F)	5.85e-01
9	64	80.0	322	7 Q29656	HLA-B*71 VARIANT (FRAGM)	5.85e-01
10	64	80.0	322	7 Q29654	HLA-B*71 (FRAGMENT)	5.85e-01
11	64	80.0	330	7 Q02945	MHC CLASS I ANTIGEN (F)	5.85e-01
12	64	80.0	338	7 Q35510	MHC CLASS I ANTIGEN (F)	5.85e-01
13	64	80.0	340	7 Q30487	MHC CLASS I HEAVY CHAI	5.85e-01
14	64	80.0	355	7 Q29853	HLA-B ALPHA-CHAIN (FRA	5.85e-01
15	64	80.0	356	7 Q30444	MHC CLASS I CAJA-G*04	5.85e-01
16	64	80.0	356	7 Q31015	MHC CLASS I SAFU-G*04	5.85e-01
17	64	80.0	357	7 Q30483	MHC CLASS I HEAVY CHAI	5.85e-01
18	64	80.0	357	7 Q31016	MHC CLASS I ANTIGEN SO	5.85e-01
19	64	80.0	357	7 Q30993	MHC CLASS I PROTEIN (F	5.85e-01
20	64	80.0	357	7 Q30899	MHC CLASS I PIP1-G*05	5.85e-01

21	64	80.0	357	7 Q30885	MHC CLASS I A (FRAGMEN	5.85e-01
22	64	80.0	359	7 Q30595	MHC CLASS I MAMU-B*01.	5.85e-01
23	64	80.0	361	7 Q46882	MHC CLASS I DLA-B*88.	5.85e-01
24	64	80.0	362	7 Q19755	MHC CLASS I ANTIGEN HL	5.85e-01
25	64	80.0	362	7 P79523	MHC CLASS I HISTOCOMPA	5.85e-01
26	64	80.0	362	7 Q29938	MHC CLASS I LYMPHOCYTE	5.85e-01
27	64	80.0	362	7 Q29938	MHC CLASS I ANTIGEN.	5.85e-01
28	64	80.0	362	7 Q29943	MHC CLASS I HLA-B 1515	5.85e-01
29	64	80.0	362	7 Q29848	HLA-B ALPHA-CHAIN.	5.85e-01
30	64	80.0	362	7 Q29848	MHC CLASS I HLA-A.	5.85e-01
31	64	80.0	362	7 Q29849	LYMPHOCYTE ANTIGEN.	5.85e-01
32	64	80.0	363	7 Q31612	MHC CLASS I HLA-B*73 CH	5.85e-01
33	64	80.0	365	7 P79603	MHC CLASS I HLA-A PROT	5.85e-01
34	64	80.0	365	7 Q30901	MHC CLASS I ALPHA CHAI	5.85e-01
35	64	80.0	365	7 Q29907	HLA-A*2404 (HLA-A24AK)	5.85e-01
36	64	80.0	365	7 Q43907	HUMAN LEUCOCYTE ANTIGE	5.85e-01
37	64	80.0	365	7 Q29747	MHC CLASS I HLA-A.	5.85e-01
38	64	80.0	366	7 Q29992	HLA CLASS I HEAVY CHAI	5.85e-01
39	64	80.0	366	7 Q29960	MHC HLA-C-ALPHA-2 CHAI	5.85e-01
40	64	80.0	366	7 Q29865	HUMAN LEUCOCYTE ANTIGE	5.85e-01
41	64	80.0	366	7 Q29921	MHC CLASS I.	5.85e-01
42	64	80.0	366	7 Q31605	ALPHA CHAIN OF MHC CLA	5.85e-01
43	64	80.0	366	7 Q29988	MHC CLASS I HLA-CW3.	5.85e-01
44	64	80.0	366	7 Q78211	HUMAN LEUCOCYTE ANTIGE	5.85e-01
45	64	80.0	366	7 Q78165	MHC CLASS I ANTIGEN.	5.85e-01

## ALIGNMENTS

RESULT 1  
 ID Q38836 PRELIMINARY; PRT; 1226 AA.  
 AC Q38836;  
 DT 01-JAN-1998 (TrEMBLrel. 05, Created)  
 DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)  
 DT 01-JAN-1998 (TrEMBLrel. 05, Last annotation update)  
 DE HYPOTHETICAL PROTEIN MJ1441.  
 GN MJ1441.  
 OS Methanococcus Jannaschii.  
 OC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;  
 OC Methanococcus.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 96337999.

RA BUIT C.J., WHITE O., OLSEN G.J., ZHOU L., FLEISCHMANN R.D.,  
 RA SUTTON G.G., BLAKE J.A., FITZGERALD L.M., CLAYTON R.A., GOCAYNE J.D.,  
 RA KERLAVAGE A.R., DOUGHERTY B.A., TOMB J.-F., ADAMS M.D., REICH C.I.,  
 RA OVERBECK R., KIRKNESS E.F., WEINSTOCK K.G., MERRICK J.M., GLODEK A.,  
 RA SCOTT J.L., GEORGEHAGEN N.S.M., WEIDMAN J.F., FUHRMANN J.L., NGUYEN D.,  
 RA UTTERBACK T.R., KELLEY J.M., PETERSON J.D., SADOW P.W., HANNA M.C.,  
 RA COTTON M.D., ROBERTS K.M., HURST M.A., KAINE B.P., BORODOVSKY M.,  
 RA KLENK H.-P., FRASER C.M., SMITH H.O., WOESE C.R., VENTER J.C.;  
 RA "Complete genome sequence of the methanogenic archaeon, Methanococcus  
 Jannaschii.";  
 RT Science 273:1058-1073(1996).  
 RL Science 273:1058-1073(1996).  
 CC -1- SIMILARITY: STRONG TO P.DENITRIFICANS COBN AND M.JANNASCHII  
 CC MJ0907.  
 DR EMBL; U67585; AAB99452.1; -.  
 DR TIGR; MJ1441; -.  
 KW Hypothetical protein.  
 SQ SEQUENCE 1226 AA; 141327 MW; 4223043D CRC32;

Query Match 85.0%; Score 68; DB 1; Length 1226;  
 Best Local Similarity 66.7%; Pred. No. 1.06e-01;  
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 845 YPENIALYW 853  
 |||:|:|:|  
 QY 1 YPAEITLYW 9

RESULT 2  
 ID P79607 PRELIMINARY; PRT; 335 AA.  
 AC P79607;



DR WPI; 97-021228/02.  
 DR N-PSDB; T59214.  
 PT Recombinant influenza haemagglutinin produced in baculovirus system  
 PT - avoids problems of growing virus in eggs and produces stable,  
 PT un-cleaved protein useful in vaccines  
 PS Example 3; Page 75-77; 107pp; English.  
 CC Recombinant influenza haemagglutinin (HA) expressed in a  
 CC baculovirus expression system in cultured insect cells, allows vaccine  
 CC production without the need to grow virus in eggs. A purer, less  
 CC allergenic product is obtained and antigen drift caused by passages  
 CC through eggs is avoided. There is no need for viral inactivation or  
 CC organic solvent extn. of viral membrane components and vaccines can be  
 CC prep'd. rapidly and cost effectively from primary sources of infection.  
 CC Recombinant HA is more stable (esp. for B strains) than HA1/HA2 complexes  
 CC and maintain correct folding during purification and storage. The present  
 CC sequence shows the N-terminal end of the HA protein for influenza  
 CC B/Panama/45/90 (sequence range 1-434).  
 SQ Sequence 585 AA;

Query Match 67.1%; Score 47; DB 1; Length 585;  
 Best Local Similarity 44.4%; Pred. NO. 1.64e+02;  
 Matches 4; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 524 FSLPTFDSL 532  
 |::| |::|  
 Qy 1 FAMPNFQTL 9

Search completed: Fri Apr 14 23:24:41 2000  
 Job time : 40 secs.

DB 243 FSLPTFDSL 251  
 QY 1 FAMPNFQTL 9

RESULT 14

ID W46786 standard; Protein; 338 AA.  
 AC W46786;  
 DE 30-JUN-1998 (first entry)  
 DT Mycobacterium tuberculosis  
 DE Exported protein; DES; antigenic protein; tuberculosis; leprosy;  
 KW Class II diiron-oxo protein family; soluble staroyl-ACP desaturase;  
 KW hybridisation probe; detection; mycobacterium species; immunoassay;  
 KW vaccine.  
 OS Mycobacterium tuberculosis.  
 PN W09804711-A2.  
 PD 05-FEB-1998.  
 PF 25-JUL-1997; IB09223.  
 PR 26-JUL-1996; US-022713.  
 PA (INSP) INST PASTEUR.  
 PI Gicquel B, Jackson M;  
 DR WPI: 98-130699/12.  
 DR N-PSDB: V16403.  
 PT DNA encoding desaturase from Mycobacterium tuberculosis - and  
 PT related probes, vectors, transforming cells, poly(peptide(s)) and  
 PT antibodies, used for detecting mycobacteria and as immunogen  
 PT Claim 13: Pages 20-21; 33pp; English.  
 CC The present sequence represents a Mycobacterium tuberculosis exported  
 CC protein designated DES, identified by using PhOA gene fusion  
 CC methodology. The des gene appears to be conserved among mycobacterial  
 CC species, and encodes an antigenic protein that is highly recognised by  
 CC human sera from both tuberculosis and leprosy patients but not by sera  
 CC from tuberculous cattle. The amino acid sequence of the DES protein  
 CC contains 2 sets of motifs that are characteristic of the active sites of  
 CC enzymes from the class II diiron-oxo protein family. Among this family,  
 CC the DES protein has significant homology to soluble staroyl-ACP  
 CC desaturases. The DNA sequence can be used to produce hybridisation  
 CC probes for detecting Mycobacterium species. The DES protein and its  
 CC fragments can be used similarly in immunoassays, and as immunogens in  
 CC vaccines. 338 AA;  
 SQ Sequence 338 AA;

Query Match 67.1%; Score 47; DB 1; Length 338;  
 Best Local Similarity 71.4%; Pred. No. 1.64e-02;  
 Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

DB 230 FQMPGFQ 236  
 QY 1 FAMPNFQ 7

RESULT 15

ID W01671 standard; Protein; 585 AA.  
 AC W01671;  
 DE 19-AUG-1997 (first entry)  
 DT Influenza B/Panama/45/90 recombinant haemagglutinin protein.  
 DE Primer: PCR; polymerase chain reaction; universal; amplify; HA;  
 KW haemagglutinin; recombinant production; baculovirus expression system;  
 KW vaccine; insect cell culture.  
 OS Synthetic.  
 FH Key  
 FH Location/Qualifiers  
 FT peptide 1..17  
 FT /label= ACPVP\_61K\_protein\_signal\_sequence  
 FT 18..568  
 FT /label= mature\_recombinant\_haemagglutinin  
 FT W09637624-A1.  
 PN 28-NOV-1996.  
 PD 26-MAY-1995; U06750.  
 PR 26-MAY-1995; WO-U06750.  
 PA (MICR-) MICROGENESIS INC.  
 PA (MGPM-) MG-PMC LLC.  
 PI Hackett CS, Smith GE, Voivovitz F, Voznesensky AI;  
 PI Wilkinson BE;

PS Disclosure; Page 50-52; 85pp; English.  
 CC The amino acid sequence of the Hepatitis B virus (HBV) polymerase (HBpol)  
 CC protein. The sequence was used to generate a series of peptides  
 CC (R70044-59) which induce cytotoxic T cell (CTL) responses against cells  
 CC infected with HBV. The HBpol peptides can be used, prophylactically as  
 CC vaccines, together with, or conjugated to, epitopes from other HBV  
 CC sequences that elicit T cell responses to HBV (see R70060-64). The  
 CC peptides can be used, particularly ex vivo, to stimulate CTL cells.  
 CC These cells can be reintroduced into patients who have chronic or acute  
 CC HBV infections or are carriers, especially in treatments to prevent  
 CC conversion from acute to chronic infections.  
 SQ Sequence 845 AA;

Query Match 68.6%; Score 48; DB 1; Length 845;  
 Best Local Similarity 66.7%; Pred. No. 1.28e+02;  
 Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 409 FAVPNQLSL 417  
 ||:|:|:|  
 QY 1 FAMPNFQTL 9

RESULT 9  
 ID W67738 standard; Protein; 2476 AA.  
 AC W67738;  
 DT 16-MAR-1999 (first entry)  
 DE Pig p105 zona pellucida-binding protein.  
 KW pig; porcine; sperm; egg-binding protein; zona pellucida; contraception;  
 KW fertilisation.  
 OS Sus scrofa.  
 PN US5851817-A.  
 PD 22-DEC-1998.  
 PF 19-JUL-1994; 276967.  
 PR 19-JUL-1994; US-276967.  
 PA (TEXA) UNIV TEXAS SYSTEM.  
 PI Garbers DL, Hardy DM;  
 DR WPI; 99-080410/07.  
 DR N-PSDB; V81446.  
 PT DNA encoding porcine sperm egg-binding protein - useful for  
 PT producing recombinant protein  
 PS Claim 1; Fig 8A-B; 47pp; English.  
 CC This sequence represents a pig sperm egg-binding protein designated  
 CC protein p105. Porcine sperm proteins which bind the zona pellucida in a  
 CC species-specific manner were isolated. Separation by gel electrophoresis  
 CC resulted in bands of proteins with molecular weights of 130, 150 and  
 CC 170 kD, under native conditions but bands of 105 and 45 kD were observed  
 CC when the proteins were separated under denaturing conditions. Compounds  
 CC which bind these proteins can be used for contraception or fertilisation.  
 SQ Sequence 2476 AA;

Query Match 68.6%; Score 48; DB 1; Length 2476;  
 Best Local Similarity 66.7%; Pred. No. 1.28e+02;  
 Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Db 1027 FDMCNFQGL 1035  
 ||:|:|:|  
 QY 1 FAMPNFQTL 9

RESULT 10  
 ID R60196 standard; Protein; 233 AA.  
 AC R60196;  
 DT 28-MAR-1995 (first entry)  
 DE Immunogenic fragment of influenza haemagglutinin (fusion protein).  
 KW Antigen; immunogen; vaccine; influenza; fusion protein; immunity;  
 KW haemagglutinin; neuraminidase; flu.  
 OS Influenza virus.  
 PN WO9417826-A.  
 PD 18-AUG-1994.  
 PF 01-FEB-1994; U01149.  
 PR 01-FEB-1993; US-013415.  
 PR 18-AUG-1993; US-108914.  
 PR 05-NOV-1993; US-149150.

PA (SMIK) SMITHKLINE BEECHAM CORP.  
 PI Dillon S, Kane J, Scott M, Shatzman A;  
 DR WPI; 94-279392/34.  
 DR N-PSDB; Q70192.  
 PT Vaccines against multi strain influenza virus infection - protect  
 PT against influenza A and B  
 PS Claim 8; Page 70-71; 151pp; English.  
 CC A vaccine comprising an immunogenic fragment of the HA2 subunit of  
 CC the influenza haemagglutinin (HA) protein from type A subtype IV and  
 CC type B IV may be used for stimulating protection in animals against  
 CC infection with influenza virus. The vaccine confers multi-strain  
 CC immunity against strains IV A and IV B. The vaccines may be  
 CC recombinantly produced, optionally as fusion proteins. In this  
 CC sequence the N-terminal 42 amino acids are derived from the influenza  
 CC NS1 protein and the remainder of the sequence comprises amino acids  
 CC 41-223 of the HA2 subunit of the BLHA2 subtype of influenza.  
 SQ Sequence 233 AA;

Query Match 67.1%; Score 47; DB 1; Length 233;  
 Best Local Similarity 44.4%; Pred. No. 1.64e+02;  
 Matches 4; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 172 FSLPTFDSL 180  
 ||:|:|:|  
 QY 1 FAMPNFQTL 9

RESULT 11  
 ID R38869 standard; Protein; 233 AA.  
 AC R38869;  
 DT 04-FEB-1994 (first entry)  
 DE Sequence of type B fusion protein NS1(1-42)HA2(41-223)  
 KW Vaccine; influenza virus; haemagglutinin subunit; HA2.  
 OS Synthetic.  
 PN WO9315763-A.  
 PD 19-AUG-1993.  
 PF 18-FEB-1993; U01451.  
 PR 18-FEB-1992; US-837773.  
 PA (SMIK) SMITHKLINE BEECHAM CORP.  
 PI Dillon SB, Scott M, Shatzman A;  
 DR WPI; 93-272565/34.  
 DR N-PSDB; Q47363.  
 PT Vaccine against Influenza A and B - contg. haemagglutinin 2  
 PT sub-unit of virus, and conferring multi-strain immunity  
 PS Claim 14; Page 58; 99pp; English.  
 CC Proteins of the invention are derived from the HA2 subunit of a  
 CC haemagglutinin (HA) protein, e.g., from a H3N2 subtype virus. Among  
 CC H3N2 subtype strains of influenza A include A/Udorn and A/Victoria  
 CC viruses. Examples are Aas 1-221 and 77-221 of a selected H3HA2  
 CC subunit. Fusion proteins are also claimed, which include a protein  
 CC derived from a H3N2 subtype virus fused in frame with, e.g., the  
 CC NS1 portion derived from a H1N1 subtype virus, A/PR/8/34 (Q47360).  
 CC The NS1 portion may comprise residues 1-42 or 1-81 of H1NS1.  
 CC Alternatively, the HA2 fragment may be fused to a portion of the  
 CC NS1 peptide derived from a selected type A virus, e.g. an H3  
 CC subtype virus (H3HA2), or a type B (BHA2) virus. The preferred type  
 CC B influenza virus is human virus strain B/Lee/40. A type B fusion  
 CC protein is NS1(1-42)HA2(41-223).  
 SQ Sequence 233 AA;

Query Match 67.1%; Score 47; DB 1; Length 233;  
 Best Local Similarity 44.4%; Pred. No. 1.64e+02;  
 Matches 4; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 172 FSLPTFDSL 180  
 ||:|:|:|  
 QY 1 FAMPNFQTL 9

RESULT 12  
 ID R60207 standard; Protein; 304 AA.  
 AC R60207;  
 DT 28-MAR-1995 (first entry)

PT response, and helper peptide - can bind to human leucocyte antigen  
 PT alleles, used to treat or prevent cancers, parasitic infections and  
 PT autoimmune disease  
 PS Claim 11; Page 39; 51pp; English.  
 CC W85138-283 represent helper T-cell peptides, which can bind to the  
 CC human leucocyte antigens (HLA) DR4w4, DR1 and DR7. The peptides  
 CC are used in the course of the invention. The specification describes  
 CC peptides that induce a cytotoxic T lymphocyte (CTL) response, and  
 CC T-helper peptides, that are used together to generate a CTL response for  
 CC the treatment or prevention of viral, fungal, bacterial or parasitic  
 CC infections (e.g. hepatitis, acquired immune deficiency syndrome or  
 CC malaria) or cancer (e.g. renal or cervical carcinoma, lymphoma, prostate  
 CC cancer or condyloma acuminatum). Helper T-cell peptides may be used  
 CC alone to induce a helper T cell response, e.g. in cases of autoimmune  
 CC disease, allograft rejection, allergy, Lyme disease, hepatitis,  
 CC post-streptococcal endocarditis, glomerulonephritis and food  
 CC hypersensitivity.  
 SQ Sequence 15 AA;

Query Match 68.6%; Score 48; DB 1; Length 15;  
 Best Local Similarity 66.7%; Pred. No. 1.28e+02;  
 Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 4 FAVPNLQSL 12  
 ||:||||:  
 Qy 1 FAMPNFQTL 9

## RESULT 6

ID W85409 standard; peptide; 15 AA.  
 AC W85409;  
 DT 16-FEB-1999 (first entry)  
 DE Helper T-cell peptide derived from POL protein.  
 KW Helper T-cell peptide; human leucocyte antigen; HLA; DR4w4; DR1;  
 KW DR7; cytotoxic T lymphocyte; CTL; hepatitis; autoimmune disease;  
 KW acquired immune deficiency syndrome; malaria; cancer;  
 KW allograft rejection; allergy; Lyme disease; hepatitis;  
 KW post-streptococcal endocarditis; glomerulonephritis;  
 KW food hypersensitivity.  
 OS Synthetic.  
 OS Hepatitis B virus.  
 PN WO9832456-A1.  
 PD 30-JUL-1998.  
 PF 23-JAN-1998; U01373.  
 PF 07-FEB-1997; US-037432.  
 PR 23-JAN-1997; US-036713.  
 PA (EPIM-) EPIMUNE INC.  
 PI Sette A, Sidney J, Southwood S;  
 PI WPI; 98-427679/36.  
 DT Composition containing peptide that induces cytotoxic T lymphocyte response, and helper peptide - can bind to human leucocyte antigen alleles, used to treat or prevent cancers, parasitic infections and autoimmune disease

PS Disclosure; Page 42; 51pp; English.  
 CC W85284-451 represent helper T-cell class II peptides, which can bind to  
 CC the human leucocyte antigens (HLA) DR4w4, DR1 and DR7. The peptides  
 CC are used in the course of the invention. The specification describes  
 CC peptides that induce a cytotoxic T lymphocyte (CTL) response, and  
 CC T-helper peptides, that are used together to generate a CTL response for  
 CC the treatment or prevention of viral, fungal, bacterial or parasitic  
 CC infections (e.g. hepatitis, acquired immune deficiency syndrome or  
 CC malaria) or cancer (e.g. renal or cervical carcinoma, lymphoma, prostate  
 CC cancer or condyloma acuminatum). Helper T-cell peptides may be used  
 CC alone to induce a helper T cell response, e.g. in cases of autoimmune  
 CC disease, allograft rejection, allergy, Lyme disease, hepatitis,  
 CC post-streptococcal endocarditis, glomerulonephritis and food  
 CC hypersensitivity.  
 SQ Sequence 15 AA;

Query Match 68.6%; Score 48; DB 1; Length 15;  
 Best Local Similarity 66.7%; Pred. No. 1.28e+02;  
 Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 2 FAVPNLQSL 10  
 ||:||||:  
 Qy 1 FAMPNFQTL 9

## RESULT 7

ID W85252 standard; peptide; 15 AA.  
 AC W85252;  
 DT 16-FEB-1999 (first entry)  
 DE Helper T-cell peptide derived from a POL protein.  
 KW Helper T-cell peptide; human leucocyte antigen; HLA; DR4w4; DR1; DR7;  
 KW cytotoxic T lymphocyte; CTL; hepatitis; autoimmune disease;  
 KW acquired immune deficiency syndrome; malaria; cancer;  
 KW allograft rejection; allergy; Lyme disease; hepatitis;  
 KW post-streptococcal endocarditis; glomerulonephritis;  
 KW food hypersensitivity.  
 OS Synthetic.  
 OS Hepatitis B virus.  
 PN WO9832456-A1.  
 PD 30-JUL-1998.  
 PF 23-JAN-1998; U01373.  
 PF 07-FEB-1997; US-037432.  
 PR 23-JAN-1997; US-036713.  
 PA (EPIM-) EPIMUNE INC.  
 PI Sette A, Sidney J, Southwood S;  
 PI WPI; 98-427679/36.  
 DT Composition containing peptide that induces cytotoxic T lymphocyte response, and helper peptide - can bind to human leucocyte antigen alleles, used to treat or prevent cancers, parasitic infections and autoimmune disease

PS Claim 11; Page 39; 51pp; English.

CC W85138-283 represent helper T-cell peptides, which can bind to the  
 CC human leucocyte antigens (HLA) DR4w4, DR1 and DR7. The peptides  
 CC are used in the course of the invention. The specification describes  
 CC peptides that induce a cytotoxic T lymphocyte (CTL) response, and  
 CC T-helper peptides, that are used together to generate a CTL response for  
 CC the treatment or prevention of viral, fungal, bacterial or parasitic  
 CC infections (e.g. hepatitis, acquired immune deficiency syndrome or  
 CC malaria) or cancer (e.g. renal or cervical carcinoma, lymphoma, prostate  
 CC cancer or condyloma acuminatum). Helper T-cell peptides may be used  
 CC alone to induce a helper T cell response, e.g. in cases of autoimmune  
 CC disease, allograft rejection, allergy, Lyme disease, hepatitis,  
 CC post-streptococcal endocarditis, glomerulonephritis and food  
 CC hypersensitivity.  
 SQ Sequence 15 AA;

Query Match 68.6%; Score 48; DB 1; Length 15;  
 Best Local Similarity 66.7%; Pred. No. 1.28e+02;  
 Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 2 FAVPNLQSL 10  
 ||:||||:  
 Qy 1 FAMPNFQTL 9

## RESULT 8

ID R70065 standard; protein; 845 AA.  
 AC R70065;  
 DT 06-OCT-1995 (first entry)  
 DE Hepatitis B virus polymerase protein.  
 KW Hepatitis B virus polymerase; cytotoxic T cell response; prophylactic;  
 KW vaccine; chronic; acute HBV infection; carrier.  
 OS Hepatitis B virus.  
 PN WO9503777-A.  
 PD 09-FEB-1995.  
 PF 01-AUG-1994; U08685.  
 PR 02-AUG-1993; US-100870.  
 PA (SCRI) SCRIPPS RES INST.  
 PI Chisari FV;  
 PI WPI; 95-082004/11.

DR New peptides inducing cytotoxic T lymphocytes to hepatitis B  
 PT virus - are regions of HB polymerase protein, for treating acute  
 PT and chronic infections

DE Cv3 consensus peptide derived immunogenic peptide #2.  
KW Immunogenic peptide; supermotif; HLA molecule; CTL response;  
KW therapeutic; diagnostic; cancer; viral infection; hepatitis B;  
KW hepatitis C.  
OS Synthetic.  
PN WO9603140-A1.  
PD 08-FEB-1996.  
PF 21-JUL-1995; US-278634.  
PR 21-JUL-1994; US-344824.  
PR 23-NOV-1994; US-344824.  
PR 30-MAY-1995; US-452843.  
PA (CYTE-) CYTEL CORP.  
PI Sette A, Sidney J.  
PI WPI: 96-116784/12.  
PT Compn. comprising immunogenic peptide with supermotif allowing more  
PT than one HLA mol. to bind - used to induce CTL response in patient  
PT and for in vivo and ex vivo therapeutic and diagnostic applications  
PS Claim 2; Page 26; 32pp; English.  
CC The sequences given in R89362-82 are immunogenic peptides which were  
CC use in the composition of the invention. The composition comprises  
CC an immunogenic peptide of 9-10 residues with a supermotif which  
CC allows binding of more than one HLA molecule. It pref. comprises  
CC two conserved residues, a first at the 2nd position from the N-  
CC terminal is Pro, and a 2nd at the C-terminal is Met. These peptides  
CC are used to induce a CTL response in a patient. They are also  
CC useful in compositions for in vivo and ex vivo therapeutic and  
CC diagnostic applications, e.g the treatment of cancer and viral  
CC infections, e.g. hepatitis B and C.  
SQ Sequence 9 AA;

Query Match 81.4%; Score 57; DB 1; Length 9;  
Best Local Similarity 88.9%; Pred. No. 1.25e+01;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 1 FAMPNEVTL 9  
QY 1 FAMPNFQTL 9

RESULT 3  
ID W71634 standard; Protein; 235 AA.  
AC W71634;  
DT 26-NOV-1998 (first entry)  
DE Omega-cyclohexane fatty acid biosynthesis enzyme #1 ORF2.  
KW Omega-cyclohexane fatty acid; biosynthesis; enzyme; detection;  
KW Alicyclobacillus acidocaldarius ATCC 27009; identification; microbe.  
OS Alicyclobacillus acidocaldarius.  
PN J10234376-A.  
PD 08-SEP-1998.  
PF 28-FEB-1997; 046570.  
PR (KIRI ) KIRIN BEVERAGE KK.  
PA WPI: 98-535030/46.  
DR N-PSDB; V58229.  
PT New nucleic acid - useful for detection and identification of genus  
PT Alicyclobacillus microorganism(s).  
PS Claim 5; Page 25; 37pp; Japanese.  
CC The present sequence represents a protein from ORF2 of a new nucleic  
CC acid which encodes enzymes which participate in the biosynthesis of  
CC Omega-cyclohexane fatty acid, where ORF2 has beta-ketoacyl (acyl  
CC carrier protein) reductase activity. The nucleic acid is isolated from  
CC Alicyclobacillus acidocaldarius. The present invention also describes  
CC primers and probes containing all or part of the nucleic acid from  
CC Alicyclobacillus acidocaldarius. The primers and probes may be used  
CC for detection and/or identification of a microorganism of genus  
CC Alicyclobacillus. The method can detect and identify Alicyclobacillus  
CC genus rapidly and easily.  
SQ Sequence 235 AA;

Query Match 70.0%; Score 49; DB 1; Length 235;  
Best Local Similarity 55.6%; Pred. No. 9.93e+01;  
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 123 FTLPNYATL 131  
QY 1 FAMPNFQTL 9

## RESULT 4

ID W80606 standard; Protein; 594 AA.  
AC W80606;  
DT 24-DEC-1998 (first entry)  
DE S. pneumoniae DNA primase.  
KW Streptococcus pneumoniae protein; recombinant; gene expression;  
KW DNA chip; virulence; antibody; infection; detection; treatment.  
OS Streptococcus pneumoniae.  
PN WO9826072-A1.  
PD 18-JUN-1998.  
PF 09-DEC-1997; U22578.  
PR 13-DEC-1996; US-036281.  
PA (ELIL ) LILLY & CO ELI.  
PI Baltz RH, Burgett SG, Dehoff BS, Hoskins JA, Jaskunas SR,  
PI Mills BJ, Norris FH, Peery RB, Rostek PK, Rostek PR,  
PI Skatrud PL, Smith MC, Solenberg PJ, Treadway PJ,  
PI Young Bellido ML;  
DR WPI: 98-348529/30.  
DR N-PSDB; V65288.  
PT Streptococcus pneumoniae nucleic acid sequences - used in DNA chips  
PT for evaluating gene expression, and identification of virulence  
PT genes  
PS Claim 3; Pages 155-158; 33pp; English.  
CC This sequence represents a Streptococcus pneumoniae DNA primase  
CC protein. The invention provides DNA sequences (V65201 to V65304) from  
CC the Streptococcus pneumoniae genome and corresponding protein sequences  
CC (W80605 to W80728). A recombinant host containing a vector comprising any  
CC of the above nucleic acids can be used for the recombinant expression of  
CC the protein sequences. The invention also provides a DNA chip having  
CC arrayed on it at least 15 base pair fragment of any one or more of these  
CC DNA sequences. The DNA chip can be used methods for evaluating gene  
CC expression in S. pneumoniae and for identifying virulence genes in  
CC S. pneumoniae. Antibodies that selectively bind to the above proteins or  
CC peptide fragments can be used to treat S. pneumoniae infection. The  
CC antibodies can also be used to detect S. pneumoniae cells.  
SQ Sequence 594 AA;

Query Match 70.0%; Score 49; DB 1; Length 594;  
Best Local Similarity 56.7%; Pred. No. 9.93e+01;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 493 FATPEFQVL 501  
QY 1 FAMPNFQTL 9

## RESULT 5

ID W85267 standard; peptide; 15 AA.  
AC W85267;  
DT 16-FEB-1999 (first entry)  
DE Helper T-cell peptide derived from a POL protein.  
KW Helper T-cell peptide; human leucocyte antigen; HLA; DR4w4; DR1; DR7;  
KW cytotoxic T lymphocyte; CTL; hepatitis; autoimmune disease;  
KW acquired immune deficiency syndrome; malaria; cancer;  
KW allograft rejection; allergy; Lyme disease; hepatitis;  
KW post-streptococcal endocarditis; glomerulonephritis;  
OS Synthetic.  
OS Hepatitis B virus.  
PN WO9832456-A1.  
PD 30-JUL-1998.  
PF 23-JAN-1998; U01373.  
PR 07-FEB-1997; US-037432.  
PR 23-JAN-1997; US-036713.  
PA (EPIM-) EPIMUNE INC.  
PI Sette A, Sidney J, Southwood S;  
DR WPI: 98-427679/36.  
PT Composition containing peptide that induces cytotoxic T lymphocyte

\*\*\*\*\*

W P S R E H (TW)

\*\*\*\*\*

Release 3.1A John F. Collins, Blocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution Rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Apr 14 23:24:01 2000; MasPar time 4.73 Seconds  
Tabular output not generated. 45.106 Million cell updates/sec

Title: >US-08-452-843-5  
Description: (1-9) from US08452843.pep  
Perfect Score: 70  
Sequence: 1 FAMPNFQTL 9

Scoring table: PAM 150  
Gap 15

Searched: 188963 seqs, 23686106 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: a-geneseq36  
l:geneseqp

Statistics: Mean 16.497; Variance 48.344; scale 0.341

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description	Pred. No.
1	70	100.0	9	R89366	Cw3 consensus peptide	3.64e-01
2	57	81.4	9	R89367	Cw3 consensus peptide	1.25e-01
3	49	70.0	235	W71634	Omega-cyclohexane fatt	9.93e+01
4	49	70.0	594	W80606	S. pneumoniae DNA prim	9.93e+01
5	48	68.6	15	W85267	Helper T-cell peptide	1.28e+02
6	48	68.6	15	W85409	Helper T-cell Class II	1.28e+02
7	48	68.6	15	W85252	Helper T-cell peptide	1.28e+02
8	48	68.6	845	R70065	Hepatitis B virus poly	1.28e+02
9	48	68.6	2476	W67738	Pig p105 zona pellucid	1.28e+02
10	47	67.1	233	R60196	Immunogenic fragment o	1.64e+02
11	47	67.1	233	R38869	Sequence of type B fus	1.64e+02
12	47	67.1	304	R60207	Immunogenic fragment o	1.64e+02
13	47	67.1	304	R60197	Immunogenic fragment o	1.64e+02
14	47	67.1	338	W46786	Mycobacterium tubercul	1.64e+02
15	47	67.1	585	W01671	Influenza B/Panama/45/	1.64e+02
16	47	67.1	585	W75443	Influenza virus B/Pana	1.64e+02
17	47	67.1	586	W75447	Influenza virus B/Harb	1.64e+02
18	47	67.1	586	W01675	Influenza B/Harbin/7/9	1.64e+02
19	47	67.1	589	W01672	Influenza B/Netherland	1.64e+02
20	47	67.1	589	W75444	Influenza virus B/Neth	1.64e+02
21	47	67.1	592	W01674	Influenza A/Shanghai/4	1.64e+02
22	47	67.1	592	W75446	Influenza virus B/Shan	1.64e+02
23	46	65.7	318	W80672	S. pneumoniae protein	2.11e+02

24	46	65.7	419	1	Y00171	Enterococcus faecalis	2.11e+02
25	46	65.7	450	1	Y00170	Enterococcus faecalis	2.11e+02
26	45	64.3	531	1	R97615	Rat N-acetylgalucosamin	2.70e+02
27	45	64.3	531	1	W24015	Human N-acetylgalucosam	2.70e+02
28	45	64.3	531	1	R48894	Human glycosyltransfer	2.70e+02
29	45	64.3	536	1	R97614	Rat N-acetylgalucosamin	2.70e+02
30	45	64.3	536	1	W24014	Rat N-acetylgalucosamin	2.70e+02
31	45	64.3	727	1	R05533	Fragment of Heymann ne	2.70e+02
32	45	64.3	4655	1	W43311	Human calcium sensor p	2.70e+02
33	45	64.3	4655	1	W43314	Human parathyroid calc	2.70e+02
34	45	64.3	4655	1	R97210	Human kidney calcium s	2.70e+02
35	45	64.3	4655	1	R97211	Human parathyroid calc	2.70e+02
36	45	64.3	4655	1	W43312	Human placental calciu	2.70e+02
37	45	64.3	4655	1	W43313	Human kidney calcium s	2.70e+02
38	45	64.3	4655	1	R97209	Human placental calciu	2.70e+02
39	45	64.3	4655	1	R97208	Human calcium sensor p	2.70e+02
40	44	62.9	211	1	W81461	Bacillus stearothermop	3.44e+02
41	44	62.9	277	1	R37312	Non-glycosylated TfPI.	3.44e+02
42	44	62.9	955	1	W31363	Cell membrane protom-A	3.44e+02
43	44	62.9	1138	1	R06461	BtPES1245 protoxin.	3.44e+02
44	43	61.4	445	1	W93956	Human beta-tubulin pro	4.39e+02
45	43	61.4	448	1	R99423	Dirofilaria immitis be	4.39e+02

## ALIGNMENTS

RESULT 1  
ID R89366 standard; peptide; 9 AA.  
AC R89366;  
DT 18-SEP-1996 (first entry)  
DE Cw3 consensus peptide derived immunogenic peptide #1.  
KW Immunogenic peptide; supermotif; HLA molecule; CTL response;  
KW therapeutic; diagnostic; cancer; viral infection; hepatitis B;  
KW hepatitis C.  
OS Synthetic.  
PN W09603140-A1.  
PD 08-FEB-1996.  
PF 21-JUL-1995; U09234.  
PR 21-JUL-1994; US-278634.  
PR 23-NOV-1994; US-344824.  
PR 30-MAY-1995; US-452843.  
PA (CYTE-) CYTEL CORP.  
PI Sette A, Sidney J.  
DR WPI; 96-116784/12.  
PT Compn. comprising immunogenic peptide with supermotif allowing more than one HLA mol. to bind - used to induce CTL response in patient and for in vivo and ex vivo therapeutic and diagnostic applications  
PS Claim 2; Page 26; 32pp; English.  
CC The sequences given in R89362-82 are immunogenic peptides which were used in the composition of the invention. The composition comprises an immunogenic peptide of 9-10 residues with a supermotif which allows binding of more than one HLA molecule. It pref. comprises two conserved residues, a first at the 2nd position from the N-terminal is Pro, and a 2nd at the C-terminal is Met. These peptides are used to induce a CTL response in a patient. They are also useful in compositions for in vivo and ex vivo therapeutic and diagnostic applications, e.g the treatment of cancer and viral infections, e.g. hepatitis B and C.  
SQ Sequence 9 AA;

Query Match 100.0%; Score 70; DB 1; Length 9;  
Best Local Similarity 100.0%; Pred. No. 3.64e-01;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 FAMPNFQTL 9  
Qy 1 FAMPNFQTL 9

RESULT 2  
ID R89367 standard; peptide; 9 AA.  
AC R89367;  
DT 18-SEP-1996 (first entry)

THIS PAGE IS BLANK (USTO)

Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 233 YPAEITLTW 241

|||||

QY 1 YPAEITLYW 9

Search completed: Fri Apr 14 23:20:14 2000  
Job time : 42 secs.



```
RESULT 14
ID 1B05_HUMAN STANDARD; PRT; 362 AA.
AC P30461;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE HLA CLASS I HISTOCOMPATIBILITY ANTIGEN, B-13 B*1301 ALPHA CHAIN
DE PRECURSOR (B13.1).
GN HLA-B OR HLAB.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Homiidae; Homo.
RN [1]
RX MEDLINE; 89235215.
RA PARHAM P., LAWOR D.A., LOMEN C.E., ENNIS P.D.;
RT "Diversity and diversification of HLA-A,B,C alleles.";
RL J. Immunol. 142:3937-3950(1989).
RN [2]
RX MEDLINE; 88152906.
RA ZEMMOUR J., ENNIS P.D., PARHAM P., DUPONT B.;
RT "Comparison of the structure of HLA-B*47 to HLA-B13 and its
RT relationship to 21-hydroxylase deficiency.";
RL Immunogenetics 27:281-287(1988).
RN [3]
RX MEDLINE; 96053518.
RA LIN L., TOKUNAGA K., NAKAJIMA F., ISHIKAWA Y., KASHIWASE K.,
RA TANAKA H., KAWATA S., SIBELTSEVA E., AKAZA T., TADOKORO K.,
RA SHIBATA Y., CHANDANAYONG D., JUJI T.;
RT "Both HLA-B*1301 and B*1302 exist in Asian populations and are
RT associated with different haplotypes.";
RL Hum. Immunol. 43:51-56(1995).
CC -1- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO
CC THE IMMUNE SYSTEM.
CC -1- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-
CC MICROGLOBULIN).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M24041; AAA59660.1;
CC DR EMBL; M19757; AAA52657.1;
CC DR EMBL; D50291; BAA08822.1;
CC DR HSSP; P30491; 1A1M.
CC DR MIM; 142830;
CC DR PROSITE; PS00290; IG_MHC; 1.
CC DR PFAM; PF00047; 1g; 1.
CC DR PFAM; PF00129; MHC_I; 1.
CC DR MHC I; Transmembrane; Glycoprotein; Signal; 3D-structure.
CC KW SIGNAL 1 24
CC CHAIN 25 362 HLA CLASS I HISTOCOMPATIBILITY ANTIGEN,
CC B-13 B*1301 ALPHA CHAIN.
CC FT DOMAIN 25 114 EXTRACELLULAR ALPHA-1.
CC FT DOMAIN 115 206 EXTRACELLULAR ALPHA-2.
CC FT DOMAIN 207 298 EXTRACELLULAR ALPHA-3.
CC FT DOMAIN 299 309 CONNECTING PEPTIDE.
CC FT TRANSMEM 310 333
CC FT DOMAIN 334 362 CYTOPLASMIC TAIL.
CC FT CARBOHYD 110 110 BY SIMILARITY.
CC FT DISULFID 125 188 BY SIMILARITY.
CC FT DISULFID 227 283 BY SIMILARITY.
CC SQ SEQUENCE 362 AA; 40474 MW; 28B67875 CRC32;

Query Match 80.0%; Score 64; DB 1; Length 362;
Best Local Similarity 88.9%; Pred. No. 5.39e-02;

Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Db 233 YPAEITLW 241
QY 1 YPAEITLW 9
|||||||
|
RESULT 15
ID 1B04_HUMAN STANDARD; PRT; 362 AA.
AC P30460;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE HLA CLASS I HISTOCOMPATIBILITY ANTIGEN, B-8 B*0801 ALPHA CHAIN
DE PRECURSOR.
GN HLA-B OR HLAB.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Homiidae; Homo.
RN [1]
RX MEDLINE; 89235215.
RA PARHAM P., LAWOR D.A., LOMEN C.E., ENNIS P.D.;
RT "Diversity and diversification of HLA-A,B,C alleles.";
RL J. Immunol. 142:3937-3950(1989).
RN [2]
RX MEDLINE; 97130420.
RA REID S.W., MCADAM S., SMITH K.J., KLENERMAN P., O'CALLAGHAN C.A.,
RA HARLOS K., JAKOBSEN B.K., MCMICHAEL A.J., BELL J.I., STUART D.I.,
RA JONES E.Y.;
RT "Antagonist HIV-1 gag peptides induce structural changes in HLA B8.";
RL J. Exp. Med. 184:2279-2286(1996).
CC -1- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO
CC THE IMMUNE SYSTEM.
CC -1- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-
CC MICROGLOBULIN).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M24036; AAA52662.1;
CC DR PDB; 1AGB; 16-JUN-97.
CC DR PDB; 1AGC; 16-JUN-97.
CC DR PDB; 1AGD; 16-JUN-97.
CC DR PDB; 1AGE; 16-JUN-97.
CC DR PDB; 1AGF; 16-JUN-97.
CC DR MIM; 142830;
CC DR PROSITE; PS00290; IG_MHC; 1.
CC DR PFAM; PF00047; 1g; 1.
CC DR PFAM; PF00129; MHC_I; 1.
CC DR MHC I; Transmembrane; Glycoprotein; Signal; 3D-structure.
CC KW SIGNAL 1 24
CC CHAIN 25 362 HLA CLASS I HISTOCOMPATIBILITY ANTIGEN,
CC B-8 B*0801 ALPHA CHAIN.
CC FT DOMAIN 25 114 EXTRACELLULAR ALPHA-1.
CC FT DOMAIN 115 206 EXTRACELLULAR ALPHA-2.
CC FT DOMAIN 207 298 EXTRACELLULAR ALPHA-3.
CC FT DOMAIN 299 309 CONNECTING PEPTIDE.
CC FT TRANSMEM 310 333
CC FT DOMAIN 334 362 CYTOPLASMIC TAIL.
CC FT CARBOHYD 110 110 BY SIMILARITY.
CC FT DISULFID 125 188 BY SIMILARITY.
CC FT DISULFID 227 283 BY SIMILARITY.
CC SQ SEQUENCE 362 AA; 40331 MW; 1467B8EB CRC32;

Query Match 80.0%; Score 64; DB 1; Length 362;
Best Local Similarity 88.9%; Pred. No. 5.39e-02;
```

```
SQ SEQUENCE 362 AA; 40478 MW; C91D06CC CRC32;
Query Match 80.0%; Score 64; DB 1; Length 362;
Best Local Similarity 88.9%; Pred. No. 5.39e-02;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 233 YPAEITLW 241
|||||||
QY 1 YPAEITLW 9

RESULT 12
ID 1B08_HUMAN STANDARD; PRT; 362 AA.
AC P30463;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 01-APR-1993 (Rel. 25, Last annotation update)
DE HLA CLASS I HISTOCOMPATIBILITY ANTIGEN, BW-65(B-14) B*1402 ALPHA CHAIN
DE PRECURSOR.
GN HLA-B OR HLAB.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 89235215.
RA PARHAM P., LAWLER D.A., LOMEN C.E., ENNIS P.D.;
RT "Diversity and diversification of HLA-A,B,C alleles.";
RL J. Immunol. 142:3937-3950(1989).
CC -1- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO
CC THE IMMUNE SYSTEM.
CC -1- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-
CC MICROGLOBULIN).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M24032; AAA59664.1; -.
CC DR HSSP; P30460; IAGB.
CC DR MM; 142830; -.
CC DR PROSITE; PS00290; IG_MHC; 1.
CC DR PFAM; PF00047; Ig; 1.
CC DR PFAM; PF00129; MHC_I; 1.
CC KW MHC I; Transmembrane; Glycoprotein; Signal.
CC SIGNAL 1 24
CC CHAIN 25 362 HLA CLASS I HISTOCOMPATIBILITY ANTIGEN,
CC FT DOMAIN 25 114 BW-65(B-14) B*1402 ALPHA CHAIN.
CC FT DOMAIN 115 206 EXTRACELLULAR ALPHA-1.
CC FT DOMAIN 207 298 EXTRACELLULAR ALPHA-2.
CC FT DOMAIN 299 309 EXTRACELLULAR ALPHA-3.
CC FT TRANSMEM 310 333 CONNECTING PEPTIDE.
CC FT DOMAIN 334 362 CYTOPLASMIC TAIL.
CC FT CARBOHYD 110 110 BY SIMILARITY.
CC FT DISULFID 125 188 BY SIMILARITY.
CC FT DISULFID 227 283 BY SIMILARITY.
CC SEQUENCE 362 AA; 40342 MW; BE68AC9E CRC32;

Query Match 80.0%; Score 64; DB 1; Length 362;
Best Local Similarity 88.9%; Pred. No. 5.39e-02;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 233 YPAEITLW 241
|||||||
QY 1 YPAEITLW 9

RESULT 13
```

```
ID 1B63_HUMAN STANDARD; PRT; 362 AA.
AC P30498;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE HLA CLASS I HISTOCOMPATIBILITY ANTIGEN, BW-78 B*7801 ALPHA CHAIN
DE PRECURSOR.
GN HLA-B OR HLAB.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 93056508.
RA MADRIGAL J.A., BELICH M.P., HILDEBRAND W.H., BENJAMIN R.J.,
RA LITTLE A.M., ZEMMOUR J., ENNIS P.D., WARD F.E., PETZL-BERLER M.L.,
RA MARTELL R.W., DU TOIT E.D., PARHAM P.;
RT "Distinctive HLA-A,B antigens of black populations formed by
RT interallelic conversion.";
RL J. Immunol. 149:3411-3415(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE: 90217537.
RA SEKIMATA M., HIRAIWA M., ANDRIEN M., DUPONT E., KARAKI S.,
RA YAMAMOTO J., KANO K., TAKIGUCHI M.;
RT "Alloantigen determinants and evolution of a novel HLA-B5 CREG antigen, HLA-B
RT SNA.";
RL J. Immunol. 144:3228-3233(1990).
CC -1- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO
CC THE IMMUNE SYSTEM.
CC -1- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-
CC MICROGLOBULIN).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X61708; CAA43877.1; -.
CC DR EMBL; M33573; AAA59644.1; -.
CC DR PIR; S16775; S16775.
CC DR HSSP; P30685; IAIN.
CC DR MM; 142830; -.
CC DR PROSITE; PS00290; IG_MHC; 1.
CC DR PFAM; PF00047; Ig; 1.
CC DR PFAM; PF00129; MHC_I; 1.
CC KW MHC I; Transmembrane; Glycoprotein; Signal.
CC SIGNAL 1 24
CC CHAIN 25 362 HLA CLASS I HISTOCOMPATIBILITY ANTIGEN,
CC FT DOMAIN 25 114 BW-78 B*7801 ALPHA CHAIN.
CC FT DOMAIN 115 206 EXTRACELLULAR ALPHA-1.
CC FT DOMAIN 207 298 EXTRACELLULAR ALPHA-2.
CC FT DOMAIN 299 308 EXTRACELLULAR ALPHA-3.
CC FT TRANSMEM 309 332 CONNECTING PEPTIDE.
CC FT DOMAIN 333 362 CYTOPLASMIC TAIL.
CC FT CARBOHYD 110 110 BY SIMILARITY.
CC FT DISULFID 125 188 BY SIMILARITY.
CC FT DISULFID 227 283 BY SIMILARITY.
CC SEQUENCE 362 AA; 40478 MW; 4023A9F5 CRC32;

Query Match 80.0%; Score 64; DB 1; Length 362;
Best Local Similarity 88.9%; Pred. No. 5.39e-02;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 233 YPAEITLW 241
|||||||
QY 1 YPAEITLW 9
```

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

```
CC -----
CC EMBL; M77775; AAA03690.1;
CC HSSP; P30685; 1A1N.
CC MIM; 142830;
CC DR PROSITE; PS00290; IG_MHC; 1.
CC DR PFAM; PF00047; Ig; 1.
CC DR PFAM; PF00129; MHC_I; 1.
CC KW MHC I; Transmembrane; Glycoprotein; Signal.
CC FT SIGNAL 1 24
CC CHAIN 25 362 HLA CLASS I HISTOCOMPATIBILITY ANTIGEN,
CC BW-56(BW-22) B*5602 ALPHA CHAIN.
CC FT DOMAIN 25 114
CC FT DOMAIN 115 206
CC FT DOMAIN 207 298
CC FT DOMAIN 299 308
CC FT TRANSMEM 309 332
CC FT DOMAIN 333 362
CC FT CARBOHYD 110 110
CC FT DISULFID 125 188
CC FT DISULFID 227 283
CC FT DISULFID 227 283
CC SQ SEQUENCE 362 AA; 40460 MW; 54608CPE CRC32;

Query Match 80.0%; Score 64; DB 1; Length 362;
Best Local Similarity 88.9%; Pred. No. 5.39e-02;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 233 YPAEITLW 241
QY 1 YPAEITLW 9

RESULT 10
ID 1B57_HUMAN STANDARD; PRT; 362 AA.
AC P30494;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DE HLA CLASS I HISTOCOMPATIBILITY ANTIGEN, BW-55(BW-22) B*5502 ALPHA
DE CHAIN PRECURSOR.
GN HLA-B OR HLAB.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 92148136.
RA HILDEBRAND W.H., MADRIGAL J.A., LITTLE A.-M., PARHAM P.;
RT "HLA-Bw22: a family of molecules with identity to HLA-B7 in the alpha
RT 1-helix.";
RL J. Immunol. 148:1155-1162(1992).
CC -!- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO
CC THE IMMUNE SYSTEM.
CC -!- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-
CC MICROGLOBULIN).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```
CC EMBL; M77775; AAA03690.1;
CC HSSP; P30685; 1A1N.
CC MIM; 142830;
CC DR PROSITE; PS00290; IG_MHC; 1.
CC DR PFAM; PF00047; Ig; 1.
CC DR PFAM; PF00129; MHC_I; 1.
CC KW MHC I; Transmembrane; Glycoprotein; Signal.
CC FT SIGNAL 1 24
CC CHAIN 25 362 HLA CLASS I HISTOCOMPATIBILITY ANTIGEN,
CC BW-56(BW-22) B*5601 ALPHA CHAIN.
CC FT DOMAIN 25 114
CC FT DOMAIN 115 206
CC FT DOMAIN 207 298
CC FT DOMAIN 299 308
CC FT TRANSMEM 309 332
CC FT DOMAIN 333 362
CC FT CARBOHYD 110 110
CC FT DISULFID 125 188
CC FT DISULFID 227 283
CC FT DISULFID 227 283
CC SQ SEQUENCE 362 AA; 40466 MW; D5BF98F0 CRC32;

Query Match 80.0%; Score 64; DB 1; Length 362;
Best Local Similarity 88.9%; Pred. No. 5.39e-02;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 233 YPAEITLW 241
QY 1 YPAEITLW 9

RESULT 11
ID 1B58_HUMAN STANDARD; PRT; 362 AA.
AC P30495;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DE HLA CLASS I HISTOCOMPATIBILITY ANTIGEN, BW-56(BW-22) B*5601 ALPHA
DE CHAIN PRECURSOR.
GN HLA-B OR HLAB.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 92148136.
RA HILDEBRAND W.H., MADRIGAL J.A., LITTLE A.-M., PARHAM P.;
RT "HLA-Bw22: a family of molecules with identity to HLA-B7 in the alpha
RT 1-helix.";
RL J. Immunol. 148:1155-1162(1992).
CC -!- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO
CC THE IMMUNE SYSTEM.
CC -!- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-
CC MICROGLOBULIN).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```
RA SARMIENTO U.M., STORR R.;
RT "Nucleotide sequence of a dog class I cDNA clone.";
RL Immunogenetics 31:400-404(1990).
CC -1- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO
CC THE IMMUNE SYSTEM.
CC -1- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-
CC MICROGLOBULIN).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M32283; AAA30865.1; -.
DR PIR; A45845; A45845.
DR HSP; P03989; ILSA.
DR PROSITE; PS00290; IG_MHC; 1.
DR PFAM; PF00047; Ig; 1.
DR PFAM; PF00129; MHC_I; 1.
DR PFAM; PF00129; MHC_I; 1.
DR MHC I; Transmembrane; Glycoprotein; Signal.
FT SIGNAL 1 24
FT CHAIN 25 362
FT FT DLA CLASS I HISTOCOMPATIBILITY ANTIGEN,
FT FT A9/A9 ALPHA CHAIN
FT FT EXTRACELLULAR ALPHA-1.
FT FT EXTRACELLULAR ALPHA-2.
FT FT EXTRACELLULAR ALPHA-3.
FT FT CONNECTING PEPTIDE.
FT FT CYTOPLASMIC TAIL.
FT FT BY SIMILARITY.
FT FT BY SIMILARITY.
FT FT POTENTIAL.
SQ SEQUENCE 362 AA; 40462 MW; D5250E8D CRC32;

Query Match 80.0%; Score 64; DB 1; Length 362;
Best Local Similarity 88.9%; Pred. No. 5.39e-02;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 234 YPAEITLW 242
| | | | | | | |
QY 1 YPAEITLW 9

RESULT 8
ID HLAF_HUMAN STANDARD; PRT; 362 AA.
AC P30511;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DE 15-JUL-1999 (Rel. 38, Last annotation update)
DE HLA CLASS I HISTOCOMPATIBILITY ANTIGEN, ALPHA CHAIN F PRECURSOR (HLA F
DE ANTIGEN) (LEUKOCYTE ANTIGEN F) (CDA12).
GN HLA-F OR HLA-5.4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 90111605.
RA GERAGHTY D.E., WEI X., ORR H.T., KOLLER B.H.;
RT "Human leukocyte antigen F (HLA-F). An expressed HLA gene composed of
RT a class I coding sequence linked to a novel transcribed repetitive
RT element.";
RL J. Exp. Med. 171:1-18(1990).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 91197869.
RA LURY D., EPSTEIN H., HOLMES N.;
RT "The human class I MHC gene HLA-F is expressed in lymphocytes.";
RL Int. Immunol. 2:531-537(1990).
CC -1- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO
```

```
CC THE IMMUNE SYSTEM.
CC -1- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-
CC MICROGLOBULIN).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X17093; CAA34947.1; -.
DR PIR; A60384; A60384.
DR PIR; J10147; J10147.
DR HSP; P03989; ILSA.
DR MIM; 143110; -.
DR PROSITE; PS00290; IG_MHC; 1.
DR PFAM; PF00047; Ig; 1.
DR PFAM; PF00129; MHC_I; 1.
DR PFAM; PF00129; MHC_I; 1.
DR MHC I; Transmembrane; Glycoprotein; Signal.
FT SIGNAL 1 21
FT CHAIN 22 362
FT FT HLA CLASS I HISTOCOMPATIBILITY ANTIGEN,
FT FT ALPHA CHAIN F.
FT FT EXTRACELLULAR ALPHA-1.
FT FT EXTRACELLULAR ALPHA-2.
FT FT EXTRACELLULAR ALPHA-3.
FT FT CONNECTING PEPTIDE.
FT FT CYTOPLASMIC TAIL.
FT FT BY SIMILARITY.
FT FT BY SIMILARITY.
FT FT BY SIMILARITY.
SQ SEQUENCE 362 AA; 40568 MW; E9B29521 CRC32;

Query Match 80.0%; Score 64; DB 1; Length 362;
Best Local Similarity 88.9%; Pred. No. 5.39e-02;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 230 YPAEITLW 238
| | | | | | | |
QY 1 YPAEITLW 9

RESULT 9
ID 1B59_HUMAN STANDARD; PRT; 362 AA.
AC P30496;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 01-APR-1993 (Rel. 25, Last annotation update)
DE HLA CLASS I HISTOCOMPATIBILITY ANTIGEN, BW-56(BW-22) B*5602 ALPHA
DE CHAIN PRECURSOR.
GN HLA-B OR HLAB.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 92148136.
RA HILDEBRAND W.H., MADRICAL J.A., LITTLE A.-M., PARHAM P.;
RT "HLA-Bw22: a family of molecules with identity to HLA-B7 in the alpha
RT 1-helix.";
RL J. Immunol. 148:1155-1162(1992).
CC -1- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO
CC THE IMMUNE SYSTEM.
CC -1- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-
CC MICROGLOBULIN).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
```

```
DR PROSITE; PS00290; IG_MHC; 1.
DR PFAM; PF00047; ig; 1.
DR PFAM; PF00129; MHC_I; 1.
KW MHC I; Transmembrane; Glycoprotein; Signal; 3D-structure.
FT SIGNAL 1 24
FT CHAIN 25 361
FT
FT DOMAIN 25 114
FT DOMAIN 115 206
FT DOMAIN 207 298
FT DOMAIN 299 308
FT TRANSMEM 309 332
FT DOMAIN 333 361
FT CARBOHYD 110 110
FT DISULFID 125 188
FT DISULFID 227 283
FT CONFLICT 206 206
FT CONFLICT 266 266
FT STRAND 27 38
FT TURN 39 41
FT TURN 42 52
FT TURN 53 54
FT STRAND 55 61
FT TURN 62 63
FT STRAND 70 71
FT HELIX 74 76
FT TURN 77 78
FT HELIX 81 108
FT TURN 109 110
FT TURN 113 114
FT STRAND 118 127
FT TURN 129 130
FT STRAND 133 142
FT TURN 143 144
FT STRAND 145 150
FT TURN 152 153
FT STRAND 157 159
FT HELIX 162 173
FT TURN 174 175
FT HELIX 176 185
FT TURN 186 186
FT HELIX 187 198
FT TURN 199 199
FT HELIX 200 203
FT TURN 204 204
FT STRAND 207 207
FT STRAND 210 217
FT STRAND 222 233
FT STRAND 238 243
FT TURN 244 245
FT STRAND 246 247
FT HELIX 249 251
FT STRAND 253 254
FT STRAND 258 259
FT STRAND 265 274
FT STRAND 275 276
FT HELIX 278 280
FT STRAND 281 286
FT TURN 288 289
FT STRAND 294 296
SQ SEQUENCE 361 AA; 40464 MW; 802130D5 CRC32;

Query Match
Best Local Similarity 88.0%; Score 64; DB 1; Length 361;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 233 YPAEITLW 241
| | | | | | |
QY 1 YPAEITLW 9

RESULT 6
ID HLAH_HUMAN STANDARD; PRT; 362 AA.

Query Match
Best Local Similarity 88.9%; Score 64; DB 1; Length 361;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 233 YPAEITLW 241
| | | | | | |
QY 1 YPAEITLW 9

RESULT 7
ID HLA9_CANFA STANDARD; PRT; 362 AA.
AC P18466;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DE DLA CLASS I HISTOCOMPATIBILITY ANTIGEN, A9/A9 ALPHA CHAIN PRECURSOR.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Carnivora; Fissipedia; Canidae; Canis.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 90316611.
```

```
AC P01893;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 15-DEC-1999 (Rel. 39, Last annotation update)
DE HLA CLASS I HISTOCOMPATIBILITY ANTIGEN, ALPHA CHAIN H PRECURSOR
DE (HLA-AR) (HLA-12.4).
GN HLA-H OR HLAH.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 82151002.
RA MALISSEN M., MALISSEN B., JORDAN B.R.;
RT "Exon/intron organization and complete nucleotide sequence of an HLA
gene."
RL Proc. Natl. Acad. Sci. U.S.A. 79:893-897(1982).
CC -!- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO
CC THE IMMUNE SYSTEM. COULD BE THE PRODUCT OF A PSEUDOGENE.
CC -!- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-
CC MICROGLOBULIN).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; J00191; AAA36218.1; ALT_INIT.
CC PIR; A02189; HLH12.
CC HSP; P03989; 1HSA.
CC MIM; 142800;
CC PROSITE; PS00290; IG_MHC; 1.
CC PFAM; PF00047; ig; 1.
CC PFAM; PF00129; MHC_I; 1.
CC MHC I; Transmembrane; Glycoprotein; Signal.
FT SIGNAL 1 24
FT CHAIN 25 362
FT
FT DOMAIN 25 114
FT DOMAIN 115 206
FT DOMAIN 207 298
FT DOMAIN 299 308
FT TRANSMEM 309 332
FT DOMAIN 333 362
FT CARBOHYD 110 110
FT DISULFID 227 283
FT SEQUENCE 362 AA; 40850 MW; 5E610F63 CRC32;

Query Match
Best Local Similarity 80.0%; Score 64; DB 1; Length 362;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 233 YPAEITLW 241
| | | | | | |
QY 1 YPAEITLW 9

RESULT 7
ID HLA9_CANFA STANDARD; PRT; 362 AA.
AC P18466;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DE DLA CLASS I HISTOCOMPATIBILITY ANTIGEN, A9/A9 ALPHA CHAIN PRECURSOR.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Carnivora; Fissipedia; Canidae; Canis.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 90316611.
```

FT CARBOHYD 104 104 BY SIMILARITY.  
SQ SEQUENCE 359 AA; 40409 MW; 55E15638 CRC32;

Query Match 80.08; Score 64; DB 1; Length 359;  
Best Local Similarity 88.9%; Pred. No. 5.39e-02;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 227 YPAEITLTW 235  
|||||||  
QY 1 YPAEITLTW 9

RESULT 4  
ID 1B01\_PANTR STANDARD; PRT; 359 AA.  
AC P13750;  
DT 01-JAN-1990 (Rel. 13, Created)  
DT 01-JAN-1990 (Rel. 13, Last sequence update)  
DT 01-APR-1993 (Rel. 25, Last annotation update)  
DE CHLA CLASS I HISTOCOMPATIBILITY ANTIGEN, B-1 ALPHA CHAIN PRECURSOR (FRAGMENT).  
OS Pan troglodytes (Chimpanzee).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Primates; Catarrhini; Homiidae; Pan.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 89030641.  
RA WAYER W.E., JONKER M., KLEIN D., IVANVI P., VAN SEVENTER G., KLEIN J.;  
RT "Nucleotide sequences of chimpanzee MHC class I alleles: evidence for trans-species mode of evolution.";  
RL EMBO J. 7:2765-2774(1988).  
RN [2]  
RP REVISIONS.  
RA WAYER W.;  
RL Submitted (FEB-1989) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO THE IMMUNE SYSTEM.  
CC -1- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-MICROGLOBULIN).  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC EMBL: X13115; CAA31507.1; -.  
DR PIR: S03537; S03537.  
DR HSSP: P03989; 1HSA.  
DR PROSITE: PS00290; IG\_MHC; 1.  
DR PFAM: PF00047; 19; 1.  
DR PFAM: PF00129; MHC\_I; 1.  
KW MHC I; Transmembrane; Glycoprotein; Signal.  
FT NON\_TER 1  
FT SIGNAL <1 20  
FT CHAIN 21 359 CHLA CLASS I HISTOCOMPATIBILITY ANTIGEN, B-1 ALPHA CHAIN.  
FT DOMAIN 21 110 EXTRACELLULAR ALPHA-1.  
FT DOMAIN 111 202 EXTRACELLULAR ALPHA-2.  
FT DOMAIN 203 294 EXTRACELLULAR ALPHA-3.  
FT DOMAIN 295 305 CONNECTING PEPTIDE.  
FT TRANSMEM 306 329 CYTOPLASMIC TAIL.  
FT DOMAIN 330 359 BY SIMILARITY.  
FT DISULFID 121 184 BY SIMILARITY.  
FT DISULFID 223 279 BY SIMILARITY.  
FT CARBOHYD 106 106 BY SIMILARITY.  
SQ SEQUENCE 359 AA; 40173 MW; 5395FFC9 CRC32;

Query Match 80.08; Score 64; DB 1; Length 359;  
Best Local Similarity 88.9%; Pred. No. 5.39e-02;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 229 YPAEITLTW 237  
|||||||  
QY 1 YPAEITLTW 9

RESULT 5  
ID 1B14\_HUMAN STANDARD; PRT; 361 AA.  
AC P03989;  
DT 23-OCT-1986 (Rel. 02, Created)  
DT 13-AUG-1987 (Rel. 05, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE HLA CLASS I HISTOCOMPATIBILITY ANTIGEN, B-27 ALPHA CHAIN PRECURSOR. HLA-B OR HLAB.  
GN Homo sapiens (Human).  
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Primates; Catarrhini; Homiidae; Homo.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 85138405.  
RA WEISS E.H., KUON W., DOERNER C., LANG M., RIETHMUELLER G.;  
RT "Organization, sequence and expression of the HLA-B27 gene: a molecular approach to analyze HLA and disease associations.";  
RL Immunobiology 170:367-380(1985).  
RN [2]  
RP SEQUENCE OF 25-361 FROM N.A.  
RX MEDLINE: 86149317.  
RA SZORTS H., RIETHMUELLER G., WEISS E., MEO T.;  
RT "Complete sequence of HLA-B27 cDNA identified through the characterization of structural markers unique to the HLA-A, -B, and -C allelic series.";  
RL Proc. Natl. Acad. Sci. U.S.A. 83:1428-1432(1986).  
RN [3]  
RP SEQUENCE OF 25-295.  
RX MEDLINE: 85226361.  
RA EZQUERRA A., BRAGADO R., VEGA M.A., STROMINGER J.L., WOODY J., LOPEZ DE CASTRO J.A.;  
RT "Primary structure of papain-solubilized human histocompatibility antigen HLA-B27.";  
RL Biochemistry 24:1733-1741(1985).  
RN [4]  
RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 25-300.  
RX MEDLINE: 92405152.  
RA MADDEN D.R., GORGA J.C., STROMINGER J.L., WILEY D.C.;  
RT "The three-dimensional structure of HLA-B27 at 2.1-A resolution suggests a general mechanism for tight peptide binding to MHC.";  
RL Cell 70:1035-1048(1992).  
RN [5]  
RP X-RAY CRYSTALLOGRAPHY.  
RX MEDLINE: 92018187.  
RA MADDEN D.R., GORGA J.C., STROMINGER J.L., WILEY D.C.;  
RT "The structure of HLA-B27 reveals nonamer self-peptides bound in an extended conformation.";  
RL Nature 353:321-325(1991).  
CC -1- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO THE IMMUNE SYSTEM.  
CC -1- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-MICROGLOBULIN).  
CC -1- DISEASE: THIS PROTEIN CORRELATES WITH THE DEVELOPMENT OF ANKYLOSING SPONDYLITIS.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC EMBL: X03945; CAA27578.1; ALT\_TERM.  
DR PIR: A25128; HLHUB2.  
DR PIR: S07441; S07441.  
DR PDB: 1HSA; 15-OCT-92.  
DR MIM: 142830; -.

```
FT DISULFID 100 163 BY SIMILARITY.
FT DISULFID 202 258 BY SIMILARITY.
FT NON_TER 273 273
SQ SEQUENCE 273 AA; 31677 MW; EEBFB366 CRC32;

Query Match 80.0%; Score 64; DB 1; Length 273;
Best Local Similarity 88.9%; Pred. No. 5.39e-02;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 208 YPAEITLW 216
QY 1 YPAEITLW 9
|||||

RESULT 2
ID HLA-E HUMAN STANDARD; PRT; 358 AA.
AC P13747;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE HLA CLASS I HISTOCOMPATIBILITY ANTIGEN, ALPHA CHAIN E E*0101/E*0102
DE PRECURSOR.
GN HLA-E OR HLA-E*6.2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
[1]
RP SEQUENCE FROM N.A. (E*0101).
RX MEDLINE; 88229102.
RA MIZUNO S., TRAPANI J.A., KOLLER B.H., DUPONT B., YANG S.Y.;
RT "Isolation and nucleotide sequence of a cDNA clone encoding a novel
RT HLA class I gene."
RL J. Immunol. 140:4024-4030(1988).
RN [2]
RP SEQUENCE FROM N.A. (E*0102).
RX MEDLINE; 88285691.
RA KOLLER B.H., GERAGHTY D.E., SHIMIZU Y., DEMARS R., ORR H.T.;
RT "HLA-E. A novel HLA class I gene expressed in resting T lymphocytes."
RL J. Immunol. 141:897-904(1988).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (2.85 ANGSTROMS) OF 22-295.
RX MEDLINE; 98325367.
RA O'CALLAGHAN C.A., TORMO J., WILLCOX B.E., BRAUD V.M., JAKOBSEN B.K.,
RA STUART D.I., MCMICHAEL A.J., BELL J.I., JONES E.Y.;
RT "Structural features impose tight peptide binding specificity in the
RT nonclassical MHC molecule HLA-E."
RL Mol. Cell 1:531-541(1998).
CC [1]
CC -!- FUNCTION: PREFERABLY BINDS TO A PEPTIDE DERIVED FROM THE SIGNAL
CC SEQUENCE OF MOST HLA-A, -B, -C AND -G MOLECULES.
CC [2]
CC -!- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-
CC MICROGLOBULIN).
CC [3]
CC -!- MISCELLANEOUS: THE SEQUENCE SHOWN IS THAT OF E*0101.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

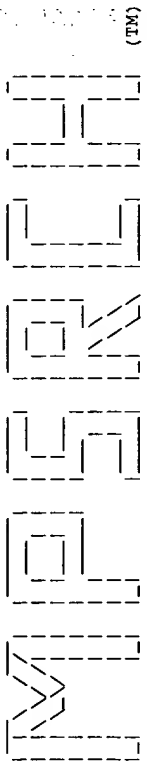
```
FT DOMAIN 22 111 ALPHA CHAIN E E*0101/E*0102.
FT DOMAIN 112 203 EXTRACELLULAR ALPHA-1.
FT DOMAIN 204 295 EXTRACELLULAR ALPHA-2.
FT DOMAIN 296 305 EXTRACELLULAR ALPHA-3.
FT TRANSMEM 306 329 CONNECTING PEPTIDE.
FT DOMAIN 330 358 CYTOPLASMIC TAIL.
FT DISULFID 122 185
FT DISULFID 224 280
FT CARBOHYD 107 107
FT VARIANT 10 10
FT VARIANT 104 104 S -> L (IN E*0102).
FT VARIANT 104 104 /FTID=VAR_004394.
FT VARIANT 104 104 G -> R (IN E*0102).
FT VARIANT 104 104 /FTID=VAR_004395.
SQ SEQUENCE 358 AA; 40130 MW; 3D79F233 CRC32;

Query Match 80.0%; Score 64; DB 1; Length 358;
Best Local Similarity 88.9%; Pred. No. 5.39e-02;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 230 YPAEITLW 238
QY 1 YPAEITLW 9
|||||

RESULT 3
ID HLA-E PONY STANDARD; PRT; 359 AA.
AC P16212;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE CLASS I HISTOCOMPATIBILITY ANTIGEN, E-1 ALPHA CHAIN PRECURSOR
DE (FRAGMENT).
OS Pongo pygmaeus (Orangutan).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Pongo.
[1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 90201944.
RA LAWLER D.A., WARREN E., WARD F.E., PARHAM P.;
RT "Comparison of class I MHC alleles in humans and apes."
RL Immunol. Rev. 113:147-185(1990).
CC [1]
CC -!- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO
CC THE IMMUNE SYSTEM.
CC [2]
CC -!- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-
CC MICROGLOBULIN).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

\*\*\*\*\*



\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Apr 14 23:19:32 2000; Maspar time 4.94 Seconds  
Tabular output not generated. 54.364 Million cell updates/sec.

Title: >US-08-452-843-4  
Description: (1-9) from US08452843.pep  
Perfect Score: 80  
Sequence: 1 YPAETLW 9

Scoring table: PAM 150  
Gap 15

Searched: 82229 seqs, 29864866 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: swiss-prot38  
1:swissprot

Statistics: Mean 25.574; Variance 33.657; scale 0.760

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	64	80.0	273	1A69_HUMAN	HLA CLASS I HISTOCOMPA	5.39e-02
2	64	80.0	358	1HAE_HUMAN	HLA CLASS I HISTOCOMPA	5.39e-02
3	64	80.0	359	1HAE_PONPY	HLA CLASS I HISTOCOMPA	5.39e-02
4	64	80.0	359	1B01_PANTR	HLA CLASS I HISTOCOMPA	5.39e-02
5	64	80.0	361	1B14_HUMAN	HLA CLASS I HISTOCOMPA	5.39e-02
6	64	80.0	362	1HLA_HUMAN	HLA CLASS I HISTOCOMPA	5.39e-02
7	64	80.0	362	1H19_CANFA	HLA CLASS I HISTOCOMPA	5.39e-02
8	64	80.0	362	1HAF_HUMAN	HLA CLASS I HISTOCOMPA	5.39e-02
9	64	80.0	362	1B59_HUMAN	HLA CLASS I HISTOCOMPA	5.39e-02
10	64	80.0	362	1B57_HUMAN	HLA CLASS I HISTOCOMPA	5.39e-02
11	64	80.0	362	1B58_HUMAN	HLA CLASS I HISTOCOMPA	5.39e-02
12	64	80.0	362	1B08_HUMAN	HLA CLASS I HISTOCOMPA	5.39e-02
13	64	80.0	362	1B63_HUMAN	HLA CLASS I HISTOCOMPA	5.39e-02
14	64	80.0	362	1B05_HUMAN	HLA CLASS I HISTOCOMPA	5.39e-02
15	64	80.0	362	1B04_HUMAN	HLA CLASS I HISTOCOMPA	5.39e-02
16	64	80.0	362	1B03_GORGO	HLA CLASS I HISTOCOMPA	5.39e-02
17	64	80.0	362	1B02_GORGO	HLA CLASS I HISTOCOMPA	5.39e-02
18	64	80.0	362	1B02_PANTR	HLA CLASS I HISTOCOMPA	5.39e-02
19	64	80.0	362	1B02_HUMAN	HLA CLASS I HISTOCOMPA	5.39e-02
20	64	80.0	362	1B11_HUMAN	HLA CLASS I HISTOCOMPA	5.39e-02
21	64	80.0	362	1B10_HUMAN	HLA CLASS I HISTOCOMPA	5.39e-02
22	64	80.0	362	1B15_HUMAN	HLA CLASS I HISTOCOMPA	5.39e-02
23	64	80.0	362	1B12_HUMAN	HLA CLASS I HISTOCOMPA	5.39e-02

24	64	80.0	362	1B13_HUMAN	HLA CLASS I HISTOCOMPA	5.39e-02
25	64	80.0	362	1B07_HUMAN	HLA CLASS I HISTOCOMPA	5.39e-02
26	64	80.0	363	1B04_GORGO	CLASS I HISTOCOMPA	5.39e-02
27	64	80.0	363	1C01_SAGOE	CLASS I HISTOCOMPA	5.39e-02
28	64	80.0	365	1C01_SAGOE	CLASS I HISTOCOMPA	5.39e-02
29	64	80.0	365	1B01_GORGO	CLASS I HISTOCOMPA	5.39e-02
30	64	80.0	365	1B01_GORGO	CLASS I HISTOCOMPA	5.39e-02
31	64	80.0	365	1A02_GORGO	CLASS I HISTOCOMPA	5.39e-02
32	64	80.0	365	1A23_HUMAN	HLA CLASS I HISTOCOMPA	5.39e-02
33	64	80.0	365	1A43_HUMAN	HLA CLASS I HISTOCOMPA	5.39e-02
34	64	80.0	365	1A36_HUMAN	HLA CLASS I HISTOCOMPA	5.39e-02
35	64	80.0	365	1A31_HUMAN	HLA CLASS I HISTOCOMPA	5.39e-02
36	64	80.0	365	1A32_HUMAN	HLA CLASS I HISTOCOMPA	5.39e-02
37	64	80.0	365	1A29_HUMAN	HLA CLASS I HISTOCOMPA	5.39e-02
38	64	80.0	365	1A30_HUMAN	HLA CLASS I HISTOCOMPA	5.39e-02
39	64	80.0	365	1A33_HUMAN	HLA CLASS I HISTOCOMPA	5.39e-02
40	64	80.0	365	1A68_HUMAN	HLA CLASS I HISTOCOMPA	5.39e-02
41	64	80.0	365	1A74_HUMAN	HLA CLASS I HISTOCOMPA	5.39e-02
42	64	80.0	366	1C02_GORGO	HLA CLASS I HISTOCOMPA	5.39e-02
43	64	80.0	366	1C01_HUMAN	HLA CLASS I HISTOCOMPA	5.39e-02
44	64	80.0	366	1C01_PANTR	HLA CLASS I HISTOCOMPA	5.39e-02
45	64	80.0	366	1C16_HUMAN	HLA CLASS I HISTOCOMPA	5.39e-02

ALIGNMENTS

RESULT ID	1A69_HUMAN	STANDARD	PRT	273 AA.
AC	PI0316;			
DT	01-MAR-1989 (Rel. 10, Last sequence update)			
DT	01-MAR-1989 (Rel. 10, Last annotation update)			
DE	HLA CLASS I HISTOCOMPATIBILITY ANTIGEN, AW-69(A-28) ALPHA CHAIN (FRAGMENT).			
GN	HLA-A OR HLAA.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OC	[1]			
RN	SEQUENCE FROM N.A. (A*6901).			
RP	MEDLINE; 86055720.			
RX	HOLMES N., PARHAM P.;			
RA	"Exon shuffling in vivo can generate novel HLA class I molecules."			
RT	EMBO J. 4:2849-2854(1985).			
RL	EMBO J. 4:2849-2854(1985).			
CC	-!- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO THE IMMUNE SYSTEM.			
CC	-!- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-MICROGLOBULIN).			
CC	-!- POLYMORPHISM: THE ONLY ALLELE OF AW-69 KNOWN IS A*6901 WHICH IS SHOWN HERE.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).			
CC	-----			
CC	EMBL; X03158; -; NOT_ANNOTATED_CDS.			
CC	EMBL; X03159; -; NOT_ANNOTATED_CDS.			
DR	PIR; B24671; HLH069.			
DR	HSSP; P01892; LAQD.			
DR	MIM; 142800; -			
DR	PROSITE; PS00290; IG_MHC; 1.			
DR	PFAM; PF00047; Ig; 1.			
DR	PFAM; PF00129; MHC_I; 1.			
KW	MHC I; Transmembrane; Glycoprotein.			
FT	NON_TER 1			
FT	DOMAIN <1 89			
FT	DOMAIN 90 180			
FT	DOMAIN 181 273			
FT	CARBOHYD 85 85			
FT	EXTRACELLULAR ALPHA-1.			
FT	EXTRACELLULAR ALPHA-2.			
FT	EXTRACELLULAR ALPHA-3.			
FT	BY SIMILARITY.			



THIS PAGE BLANK (USPTO)

```
RESULT 14
ENTRY HLHUB4 #type fragment
TITLE MHC class I histocompatibility antigen HLA-B*44 alpha chain
        precursor - human (fragment)
ORGANISM #formal_name Homo sapiens #common_name man
DATE 31-Dec-1988 #sequence_revision 31-Dec-1988 #text_change
        20-Mar-1998
ACCESSIONS A25295
REFERENCE Kottmann, A.H.; Seemann, G.H.A.; Gressow, H.D.; Roos, M.H.
        Immunogenetics (1986) 23:396-400
        #journal DNA sequence of the coding region of the HLA-B*44 gene.
        #cross-references MUID:86249389
        #accession A25295
        #molecule_type mRNA
        #residues 1-359 #label KOT
        ##cross-references GB:M15470; NID:g187680; PID:g386883
GENETICS
        #gene GDB:HLA-B
        #cross-references GDB:120048; OMIM:142830
        #map_position 6p21.3-6p21.3
CLASSIFICATION #superfamily class I histocompatibility antigen;
        immunoglobulin homology
KEYWORDS duplication; glycoprotein; heterodimer; transmembrane
        protein; transplantation antigen
FEATURE
1-21 #domain signal sequence (fragment) #status predicted
        #label SIG\
22-359 #product class I histocompatibility antigen HLA-B*44
        alpha chain #status predicted #label MAT\
22-304 #domain extracellular #status predicted #label EXT\
22-111 #domain alpha-1 #label EX1\
112-203 #domain alpha-2 #label EX2\
217-282 #domain immunoglobulin homology #label IMM\
305-328 #domain transmembrane #status predicted #label TM\
329-359 #domain intracellular #status predicted #label INT\
107 #binding_site carbohydrate (Asn) (covalent) #status
        predicted
SUMMARY #length 359 #checksum 3910
Query Match 80.0%; Score 64; DB 1; Length 359;
Best Local Similarity 88.9%; Pred. No. 2.64e-01;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 230 YPAEITLTW 238
      |||||
QY 1 YPAEITLYW 9

RESULT 15
ENTRY S25415 #type complete
TITLE class I histocompatibility antigen HLA-B*4403 alpha chain -
        human
ORGANISM #formal_name Homo sapiens #common_name man
DATE 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change
        24-Sep-1998
ACCESSIONS S25415
REFERENCE Fleischhauer, K.; Kernan, N.A.; Dupont, B.; Yang, S.Y.
        Tissue Antigens (1991) 37:133-137
        #journal The two major subtypes of HLA-B*44 differ for a single amino
        acid in codon 156.
        #cross-references MUID:91335451
        #accession S25415
        #status preliminary
        #molecule_type mRNA
        #residues 1-362 #label FILE
        ##cross-references EMBL:X64366; NID:g32178; PID:g32179
GENETICS
        #gene GDB:HLA-B
        #cross-references GDB:120048; OMIM:142830
        #map_position 6p21.3-6p21.3
CLASSIFICATION #superfamily class I histocompatibility antigen;
```

```
KEYWORDS immunoglobulin homology
FEATURE transmembrane protein
220-285 #domain immunoglobulin homology #label IMM
SUMMARY #length 362 #molecular-weight 40479 #checksum 690
Query Match 80.0%; Score 64; DB 2; Length 362;
Best Local Similarity 88.9%; Pred. No. 2.64e-01;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 233 YPAEITLTW 241
      |||||
QY 1 YPAEITLYW 9

Search completed: Fri Apr 14 23:19:13 2000
Job time : 13 secs.
```

```

immunoglobulin homology
glycoprotein; transmembrane protein
KEYWORDS
1-16 #domain signal sequence #status predicted #label SIG\
17-354 #product class I histocompatibility antigen HLA-B-1504
#status predicted #label MAT\
17-299 #domain extracellular #status predicted #label EXT\
107-198 #domain alpha-2 #status predicted #label EX2\
212-277 #domain immunoglobulin homology #label IMM\
300-323 #domain transmembrane #status predicted #label TMM\
324-354 #domain intracellular #status predicted #label INT\
102 #binding-site carbohydrate (Asn) (covalent) #status
predicted\
117-180,219-275 #disulfide_bonds #status predicted
SUMMARY #length 354 #checksum 5322
Query Match 80.0%; Score 64; DB 2; Length 354;
Best Local Similarity 88.9%; Pred. No. 2.64e-01;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Db 225 YPAEITLW 233
| | | | | |
QY 1 YPAEITLW 9

RESULT 11
ENTRY HLH12 #type complete
TITLE MHC class I histocompatibility antigen alpha chain - cotton-top
ORGANISM tamarin
DATE #formal_name Saginus oedipus #common_name cotton-top tamarin
25-Feb-1994 #sequence_revision 26-May-1995 #text_change
S11141
AUTHORS #authors
S10934
#journal
#title
#cross-references MUID:90309971
#accession
#status preliminary; nucleic acid sequence not shown; not
#molecule_type mRNA
#residues 1-357 #label WAT
CLASSIFICATION #superfamily class I histocompatibility antigen;
immunoglobulin homology
FEATURE #domain immunoglobulin homology #label IMM
212-277 #length 357 #molecular-weight 40116 #checksum 4093
SUMMARY
Query Match 80.0%; Score 64; DB 2; Length 357;
Best Local Similarity 88.9%; Pred. No. 2.64e-01;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Db 225 YPAEITLW 233
| | | | | |
QY 1 YPAEITLW 9

RESULT 12
ENTRY HLH12 #type complete
TITLE MHC class I histocompatibility antigen alpha chain - cotton-top
ORGANISM tamarin
DATE #formal_name Saginus oedipus #common_name cotton-top tamarin
25-Feb-1994 #sequence_revision 26-May-1995 #text_change
S11139
AUTHORS #authors
S10934
#journal
#title
#cross-references MUID:90309971
#accession
#status preliminary; nucleic acid sequence not shown; not
#molecule_type mRNA
#residues 1-357 #label WAT
CLASSIFICATION #superfamily class I histocompatibility antigen;
immunoglobulin homology
FEATURE #domain immunoglobulin homology #label IMM
212-277 #length 357 #molecular-weight 40116 #checksum 4093
SUMMARY
Query Match 80.0%; Score 64; DB 2; Length 357;
Best Local Similarity 88.9%; Pred. No. 2.64e-01;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Db 225 YPAEITLW 233
| | | | | |
QY 1 YPAEITLW 9

RESULT 13
ENTRY HLH12 #type complete
TITLE MHC class I histocompatibility antigen HLA alpha chain
precursor (clone pHLA 12.4) - human
ORGANISM #formal_name Homo sapiens #common_name man
05-Apr-1993 #sequence_revision 05-Apr-1983 #text_change
20-Mar-1998
A02189
AUTHORS #authors
Malissen, M.; Malissen, B.; Jordan, B.R.
Proc. Natl. Acad. Sci. U.S.A. (1982) 79:893-897
#journal
#title
#cross-references MUID:82151002
#accession
#molecule_type DNA
#residues 1-359 #label MAL
#cross-references GB:J00191; GB:Y00526; NID:G187600; PID:G386873
COMMENT The seven exons correspond approximately to the domain structure of
this chain.
GENETICS
#map_position 6p21.3
#introns 22/1; 112/1; 204/1; 296/1; 335/1; 346/1
CLASSIFICATION #superfamily class I histocompatibility antigen;
immunoglobulin homology
KEYWORDS duplication; glycoprotein; heterodimer; transmembrane
protein; transplantation antigen
FEATURE #domain signal sequence #status predicted #label SIG\
1-21 #product class I histocompatibility antigen HLA alpha
chain #status predicted #label MAT\
22-359 #domain extracellular #status predicted #label EXT\
22-304 #domain alpha-1 #label EX1\
22-111 #domain alpha-2 #label EX2\
212-203 #domain immunoglobulin homology #label IMM\
217-282 #domain transmembrane #status predicted #label TMM\
305-329 #domain intracellular #status predicted #label INT\
335-359 #binding-site carbohydrate (Asn) (covalent) #status
predicted\
224-280 #disulfide_bonds #status predicted
SUMMARY #length 359 #molecular-weight 40548 #checksum 2195
Query Match 80.0%; Score 64; DB 1; Length 359;
Best Local Similarity 88.9%; Pred. No. 2.64e-01;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Db 230 YPAEITLW 238
| | | | | |
QY 1 YPAEITLW 9

```

```

from ancestral homologues of human non-classical genes.
#cross-references MUID:90309971
#accession S11139
#status preliminary; nucleic acid sequence not shown; not
#molecule_type mRNA
#residues 1-357 #label WAT
CLASSIFICATION #superfamily class I histocompatibility antigen;
immunoglobulin homology
FEATURE #domain immunoglobulin homology #label IMM
212-277 #length 357 #molecular-weight 40193 #checksum 7731
SUMMARY
Query Match 80.0%; Score 64; DB 2; Length 357;
Best Local Similarity 88.9%; Pred. No. 2.64e-01;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Db 225 YPAEITLW 233
| | | | | |
QY 1 YPAEITLW 9

RESULT 13
ENTRY HLH12 #type complete
TITLE MHC class I histocompatibility antigen HLA alpha chain
precursor (clone pHLA 12.4) - human
ORGANISM #formal_name Homo sapiens #common_name man
05-Apr-1993 #sequence_revision 05-Apr-1983 #text_change
20-Mar-1998
A02189
AUTHORS #authors
Malissen, M.; Malissen, B.; Jordan, B.R.
Proc. Natl. Acad. Sci. U.S.A. (1982) 79:893-897
#journal
#title
#cross-references MUID:82151002
#accession
#molecule_type DNA
#residues 1-359 #label MAL
#cross-references GB:J00191; GB:Y00526; NID:G187600; PID:G386873
COMMENT The seven exons correspond approximately to the domain structure of
this chain.
GENETICS
#map_position 6p21.3
#introns 22/1; 112/1; 204/1; 296/1; 335/1; 346/1
CLASSIFICATION #superfamily class I histocompatibility antigen;
immunoglobulin homology
KEYWORDS duplication; glycoprotein; heterodimer; transmembrane
protein; transplantation antigen
FEATURE #domain signal sequence #status predicted #label SIG\
1-21 #product class I histocompatibility antigen HLA alpha
chain #status predicted #label MAT\
22-359 #domain extracellular #status predicted #label EXT\
22-304 #domain alpha-1 #label EX1\
22-111 #domain alpha-2 #label EX2\
212-203 #domain immunoglobulin homology #label IMM\
217-282 #domain transmembrane #status predicted #label TMM\
305-329 #domain intracellular #status predicted #label INT\
335-359 #binding-site carbohydrate (Asn) (covalent) #status
predicted\
224-280 #disulfide_bonds #status predicted
SUMMARY #length 359 #molecular-weight 40548 #checksum 2195
Query Match 80.0%; Score 64; DB 1; Length 359;
Best Local Similarity 88.9%; Pred. No. 2.64e-01;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Db 230 YPAEITLW 238
| | | | | |
QY 1 YPAEITLW 9

```

```
GENETICS
#gene
#map_position 621.3-6p21.3
#cross-references GDB:119311; OMIM:142840
CLASSIFICATION #superfamily class I histocompatibility antigen;
immunoglobulin homology
KEYWORDS glycoprotein; heterodimer; transmembrane protein;
transplantation antigen
FEATURE
196-261 #domain immunoglobulin homology #label IMM\
86 #binding_site carbohydrate (Asn) (covalent) #status
predicted
SUMMARY #length 342 #molecular-weight 38082 #checksum 7418

Query Match 80.0%; Score 64; DB 1; Length 342;
Best Local Similarity 88.9%; Pred. No. 2.64e-01;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 209 YPAEITLW 217
QY 1 YPAEITLW 9

RESULT 7
ENTRY #type fragment
TITLE MHC HLA B71 - human (fragment)
ORGANISM #formal_name Homo sapiens #common_name man
DATE 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change
29-Aug-1997
ACCESSIONS I54308
REFERENCE Rodriguez, S.G.; Johnson, A.H.; Hurley, C.K.
#authors Hum. Immunol. (1993) 37:192-194
#journal Molecular characterization of HLA-B71 from an African
#title American individual.
#cross-references MUID:94064392
#accession I54308
#status preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues 1-350 #label RES
#cross-references GB:L07950; NID:9307236; PID:g307237

GENETICS
#gene
#map_position 621.3-6p21.3
#cross-references GDB:120048; OMIM:142830
CLASSIFICATION #superfamily class I histocompatibility antigen;
immunoglobulin homology
SUMMARY #length 350 #checksum 7005

Query Match 80.0%; Score 64; DB 2; Length 350;
Best Local Similarity 88.9%; Pred. No. 2.64e-01;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 233 YPAEITLW 241
QY 1 YPAEITLW 9

RESULT 8
ENTRY #type fragment
TITLE class I histocompatibility antigen - pygmy chimpanzee
(fragment)
ORGANISM #formal_name Pan paniscus #common_name pygmy chimpanzee,
bonobo
DATE 31-May-1996 #sequence_revision 31-May-1996 #text_change
16-Feb-1997
ACCESSIONS I59308
REFERENCE McAdam, S.N.; Boyson, J.E.; Liu, X.; Garber, T.L.; Hughes,
A.L.; Bontrop, R.E.; Watkins, D.I.
#authors Proc. Natl. Acad. Sci. U.S.A. (1994) 91:5893-5897
#journal A uniquely high level of recombination at the HLA-B locus.
#title
#cross-references MUID:94286544

GENETICS
#gene
#map_position 621.3-6p21.3
#cross-references GDB:119311; OMIM:142840
CLASSIFICATION #superfamily class I histocompatibility antigen;
immunoglobulin homology
SUMMARY #length 354 #checksum 3211

Query Match 80.0%; Score 64; DB 2; Length 354;
Best Local Similarity 88.9%; Pred. No. 2.64e-01;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 225 YPAEITLW 233
QY 1 YPAEITLW 9

RESULT 9
ENTRY #type fragment
TITLE class I histocompatibility antigen - pygmy chimpanzee
(fragment)
ORGANISM #formal_name Pan paniscus #common_name pygmy chimpanzee,
bonobo
DATE 24-May-1996 #sequence_revision 24-May-1996 #text_change
16-Feb-1997
ACCESSIONS I80166
REFERENCE McAdam, S.N.; Boyson, J.E.; Liu, X.; Garber, T.L.; Hughes,
A.L.; Bontrop, R.E.; Watkins, D.I.
#authors Proc. Natl. Acad. Sci. U.S.A. (1994) 91:5893-5897
#journal A uniquely high level of recombination at the HLA-B locus.
#title
#cross-references MUID:94286544
#accession I80166
#status preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues 1-354 #label RES
#cross-references EMBL:U05577; NID:g454771; PID:g454772
CLASSIFICATION #superfamily class I histocompatibility antigen;
immunoglobulin homology
SUMMARY #length 354 #checksum 3443

Query Match 80.0%; Score 64; DB 2; Length 354;
Best Local Similarity 88.9%; Pred. No. 2.64e-01;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 225 YPAEITLW 233
QY 1 YPAEITLW 9

RESULT 10
ENTRY #type fragment
TITLE class I histocompatibility antigen HLA-B-1504 precursor -
human (fragment)
ORGANISM #formal_name Homo sapiens #common_name man
DATE 22-Nov-1993 #sequence_revision 18-Jul-1997 #text_change
07-Nov-1997
ACCESSIONS S24433
REFERENCE Watkins, D.I.; McAdam, S.N.; Liu, X.; Strang, C.R.; Milford,
E.L.; Levine, C.G.; Garber, T.L.; Dogan, A.L.; Lord, C.I.;
Ghim, S.H.; Troup, G.M.; Hughes, A.L.; Letvin, N.L.
#authors Nature (1992) 357:329-333
#journal New recombinant HLA-B alleles in a tribe of South American
#title Amerindians indicate rapid evolution of MHC class I loci.
#cross-references MUID:92269956
#accession S24433
#molecule_type mRNA
#residues 1-354 #label WAT
#gene HLA-B-1504
CLASSIFICATION #superfamily class I histocompatibility antigen;
```

```
ORGANISM      #formal_name Homo sapiens #common_name man
DATE          07-Jun-1996 #sequence_revision 07-Jun-1996 #text_change
ACCESSIONS    I54449
REFERENCE      I54449
#authors      Arnott, D.; Lillie, J.W.; Auffray, C.; Kappes, D.; Strominger,
              J.L.
#journal      Immunogenetics (1984) 20:237-252
#title        Inter-locus and intra-allelic polymorphisms of HLA class I
              antigen gene mRNA.
#cross-references MIM:84287690
#accession    I54449
#status       preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues     1-235 #label RES
#cross-references GB:M27539; NID:gl87731; PID:g386889
CLASSIFICATION #superfamily class I histocompatibility antigen;
               immunoglobulin homology
KEYWORDS       surface antigen
SUMMARY        #length 235 #checksum 9495

Query Match      80.0%; Score 64; DB 2; Length 235;
Best Local Similarity 88.9%; Pred. No. 2.64e-01;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 103 YPAEITLW 111
   |||||
QY 1 YPAEITLW 9

RESULT 3
ENTRY   #type fragment
TITLE   MHC class I - chimpanzee (fragment)
ORGANISM #formal_name Pan troglodytes #common_name chimpanzee
DATE     02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change
09-May-1997
ACCESSIONS I36958
REFERENCE   I36958
#authors    Parham, P.; Lawlor, D.A.; Lomen, C.E.; Ennis, P.D.
#journal    J. Immunol. (1989) 142:3937-3950.
#title      Diversity and diversification of HLA-A,B,C alleles.
#cross-references MIM:89235215
#accession   I36958
#status      preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues    1-313 #label RES
#cross-references GB:M24046; NID:gl76816; PID:gl76817
CLASSIFICATION #superfamily class I histocompatibility antigen;
               immunoglobulin homology
SUMMARY        #length 313 #checksum 5311

Query Match      80.0%; Score 64; DB 2; Length 313;
Best Local Similarity 88.9%; Pred. No. 2.64e-01;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 180 YPAEITLW 188
   |||||
QY 1 YPAEITLW 9

RESULT 4
ENTRY   #type fragment
TITLE   MHC class I HLA-Cx52 - human (fragment)
ORGANISM #formal_name Homo sapiens #common_name man
DATE     01-Nov-1996 #sequence_revision 01-Nov-1996 #text_change
29-Aug-1997
ACCESSIONS I54449
REFERENCE   I54449
#authors    Takata, H.; Inoko, H.; Ando, A.; Haranaka, M.; Watanabe, B.;
              Tsuji, K.; Iri, H.
#journal    Immunogenetics (1988) 28:265-270
#title      Cloning and analysis of HLA class I cDNA encoding a new HLA-C
              specificity Cx52.
```

```
#cross-references MIM:88330144
#accession    I54449
#status       preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues     1-325 #label RES
#cross-references GB:M1963; NID:gl88540; PID:gl88541
GENETICS
#gene         GDB:HLA-C
#cross-references GDB:119311; OMIM:142840
#map_position 6p21.3-6p21.3
CLASSIFICATION #superfamily class I histocompatibility antigen;
               immunoglobulin homology
SUMMARY        #length 325 #checksum 46

Query Match      80.0%; Score 64; DB 2; Length 325;
Best Local Similarity 88.9%; Pred. No. 2.64e-01;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 192 YPAEITLW 200
   |||||
QY 1 YPAEITLW 9

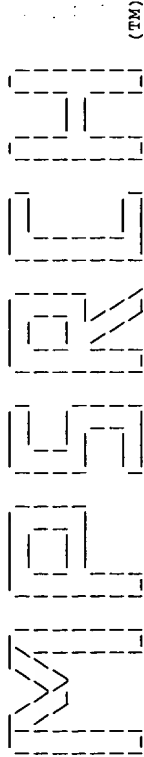
RESULT 5
ENTRY   #type fragment
TITLE   MHC HLA-B27-HS - human (fragment)
ORGANISM #formal_name Homo sapiens #common_name man
DATE     02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change
16-Feb-1997
ACCESSIONS I56116
REFERENCE   I56116
#authors    Choo, S.Y.; Fan, L.A.; Hansen, J.A.
#journal    J. Immunol. (1991) 147:174-180
#title      A novel HLA-B27 allele maps B27 allospecificity to the region
              around position 70 in the alpha 1 domain.
#cross-references MIM:91288545
#accession   I56116
#status      preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues    1-338 #label RES
#cross-references GB:M62852; NID:gl87760; PID:gl87761
CLASSIFICATION #superfamily class I histocompatibility antigen;
               immunoglobulin homology
SUMMARY        #length 338 #checksum 3677

Query Match      80.0%; Score 64; DB 2; Length 338;
Best Local Similarity 88.9%; Pred. No. 2.64e-01;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 209 YPAEITLW 217
   |||||
QY 1 YPAEITLW 9

RESULT 6
ENTRY   #type complete
TITLE   MHC class I histocompatibility antigen HLA-C4 alpha chain -
              human
ORGANISM #formal_name Homo sapiens #common_name man
DATE     31-Dec-1988 #sequence_revision 31-Dec-1988 #text_change
20-Mar-1998
ACCESSIONS A24512
REFERENCE   A24512
#authors    Davidson, W.F.; Kress, M.; Khoury, G.; Jay, G.
#journal    J. Biol. Chem. (1985) 260:13414-13423
#title      Comparison of HLA class I gene sequences. Derivation of
              locus-specific oligonucleotide probes specific for HLA-A,
              HLA-B, and HLA-C genes.
#cross-references MIM:86033791
#accession   A24512
#molecule_type DNA
#residues    1-342 #label DAV
#cross-references GB:M1886; NID:gl84173; PID:g386777
```

\*\*\*\*\*



(TM)

\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPErch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Fri Apr 14 23:19:00 2000; MasPar time 3.65 Seconds  
Tabular output not generated. 98.673 Million cell updates/sec

Title: >US-08-452-843-4  
Description: (1-9) from US08452843.pep  
Perfect Score: 80  
Sequence: 1 YPAEITLYW 9

Scoring table: PAM 150  
Gap 15

Searched: 122810 seqs, 40068593 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: p1r62  
1:p1r1 2:p1r2 3:p1r3 4:p1r4

Statistics: Mean 25.125; Variance 37.800; scale 0.665

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	68	85.0	1226	H64479	protoporphyrin IX mag	4.64e-02
2	64	80.0	235	I68700	MHC HLA-A cell surfac	2.64e-01
3	64	80.0	313	I36958	MHC ChIA chain - chim	2.64e-01
4	64	80.0	325	I54449	MHC class I HLA-Cx52	2.64e-01
5	64	80.0	338	I56116	MHC HLA-B27-HS - huma	2.64e-01
6	64	80.0	342	HLHUC4	MHC class I histocomp	2.64e-01
7	64	80.0	350	I54308	MHC HLA B71 - human	2.64e-01
8	64	80.0	354	I59308	class I histocompatib	2.64e-01
9	64	80.0	354	I80166	class I histocompatib	2.64e-01
10	64	80.0	354	S24433	class I histocompatib	2.64e-01
11	64	80.0	357	S11141	class I histocompatib	2.64e-01
12	64	80.0	357	S11139	class I histocompatib	2.64e-01
13	64	80.0	359	HLHUB12	MHC class I histocomp	2.64e-01
14	64	80.0	359	HLHUB4	MHC class I histocomp	2.64e-01
15	64	80.0	362	S25415	class I histocompatib	2.64e-01
16	64	80.0	362	I37522	MHC class I histocomp	2.64e-01
17	64	80.0	362	JH0540	class I histocompatib	2.64e-01
18	64	80.0	362	HLHUB8	MHC class I histocomp	2.64e-01
19	64	80.0	362	I37519	MHC class I histocomp	2.64e-01
20	64	80.0	362	I62044	MHC class I histocomp	2.64e-01
21	64	80.0	362	JH0292	class I histocompatib	2.64e-01
22	64	80.0	362	I37492	HLA-B alpha-chain - h	2.64e-01
23	64	80.0	362	I68724	MHC class I histocomp	2.64e-01

24	64	80.0	362	2	S60601	HLA-BPOT (classI) pro	2.64e-01
25	64	80.0	362	2	S52486	HLA-B protein alpha c	2.64e-01
26	64	80.0	362	2	I62045	gene HLA B-1517 prote	2.64e-01
27	64	80.0	362	2	I61864	MHC HLA-Bw41 chain -	2.64e-01
28	64	80.0	362	2	I62043	MHC HLA-B cell surfac	2.64e-01
29	64	80.0	362	2	I81233	lymphocyte antigen -	2.64e-01
30	64	80.0	363	2	S07113	class I histocompatib	2.64e-01
31	64	80.0	364	2	I72217	class I histocompatib	2.64e-01
32	64	80.0	364	2	A35997	MHC class I histocomp	2.64e-01
33	64	80.0	365	2	I38518	HLA-A*0102 allele - h	2.64e-01
34	64	80.0	365	2	JH0544	class I histocompatib	2.64e-01
35	64	80.0	365	2	I38441	gene HLA-A-6802 prote	2.64e-01
36	64	80.0	365	2	I38519	MHC class I histocomp	2.64e-01
37	64	80.0	365	2	I37470	HLA-A*0210 - human	2.64e-01
38	64	80.0	365	2	I83063	All.2 - human	2.64e-01
39	64	80.0	365	2	I37476	MHC class I histocomp	2.64e-01
40	64	80.0	365	2	I58039	HLA-A30.3 precursor -	2.64e-01
41	64	80.0	365	2	A45847	MHC class I histocomp	2.64e-01
42	64	80.0	366	2	JH0546	class I histocompatib	2.64e-01
43	64	80.0	366	2	JH0545	class I histocompatib	2.64e-01
44	64	80.0	366	2	I37526	MHC class I histocomp	2.64e-01
45	64	80.0	381	2	S35940	class I histocompatib	2.64e-01

ALIGNMENTS

RESULT 1

ENTRY

TITLE

ORGANISM

DATE

ACCESSIONS

REFERENCE

#authors

H64479 #type complete  
protoporphyrin IX magnesium chelate (EC 4.99.1.-) homolog -  
Methanococcus jannaschii  
#formal\_name Methanococcus jannaschii  
13-Sep-1996 #sequence\_revision 13-Sep-1996 #text\_change  
21-Aug-1998  
H64479  
R64300  
Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake, J.A.; Fitzgerald, L.M.; Clayton, R.A.; Gocayne, J.D.; Kerlavage, A.R.; Dougherty, B.A.; Tomb, J.F.; Adams, M.D.; Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.; Scott, J.L.; Geoghegan, N.S.M.; Weidman, J.F.; Fuhrmann, J.L.; Nguyen, D.; Utterback, T.R.; Kelley, J.M.; Peterson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.; Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese, C.R.; Venter, J.C.  
Science (1996) 273:1058-1073  
Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii.  
#cross-references MUID:96337999  
#accession H64479  
#status preliminary; nucleic acid sequence not shown;  
translation not shown  
#molecule\_type DNA  
#residues 1-1226 #label BUL  
#cross-references GB:U67585; GB:L77117; NID:gl592088; PID:gl500323; TIGR:MJ1441

#journal

#title

#cross-references

#accession

#status

#molecule\_type

#residues

#cross-references

GENETICS

#map\_position

KEYWORDS

SUMMARY

Query Match

Best Local Similarity

Matches

Db

Qy

RESULT

ENTRY

TITLE

Score 68; DB 2; Length 1226;

Best Local Similarity 66.7%; Pred. No. 4.64e-02;

Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

845 YPENTLYW 853

1 YPAEITLYW 9

I68700 #type fragment

MHC HLA-A cell surface antigen - human (fragment)

THIS PAGE BLANK (USPTO)

CC beta-2 microglobulin via a flexible linker. The method is also used to  
 CC detect and quantify tumour-specific T-cells and to generate CTC for  
 CC specific killing of tumour cells (solid tumours, leukaemia or lymphoma)  
 CC by injection into a human or animal, but also for treating viral  
 CC infections. 412 AA;  
 SQ Sequence 412 AA;

Query Match 80.0%; Score 64; DB 1; Length 412;  
 Best Local Similarity 88.9%; Pred. No. 7.35e+00;  
 Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 233 YPAEITLW 241  
 QY 1 YPAEITLYW 9  
 |||:|||||

RESULT 14  
 ID R69622 standard; peptide; 25 AA.  
 AC R69622;  
 DT 29-AUG-1995 (first entry)  
 DE MHC-I peptide Db-(197-221).  
 KW MHC class I; major histocompatibility complex; insulin receptor;  
 OS diabetics; glucose uptake; adipocyte.  
 OS Synthetic.  
 PN US538588-A.  
 PD 31-JAN-1995.  
 PF 20-MAR-1987; 028241.  
 PR 20-MAR-1987; US-028241.  
 PR 14-MAR-1989; US-323565.  
 PR 01-FEB-1991; US-649471.  
 PR 03-MAY-1993; US-057184.  
 PA (REGC ) UNIV CALIFORNIA.  
 PI Goodenow RS, Olsson L;  
 DR WPI; 95-081582/11.

PT Modulating response of cellular insulin receptor to ligand -  
 PT using peptide deriv. from MHC class I antigen, partic. to  
 PT potentiate effect of insulin for treating diabetes  
 PS Disclosure: Column 18; 15pp; English.  
 CC Response of an insulin receptor (IR) to a ligand is modulated by  
 CC contacting mammalian cells having IR on the surface with peptides  
 CC derived from MHC class I antigen. Peptide Dk-(61-85) (R69619),  
 CC from the alpha-1 domain of MHC-I, boosted glucose uptake 5-6 fold  
 CC over basal levels in rat adipocytes, when administered at 30 uM.  
 CC Peptide Db-(191-221), from the alpha-3 region, had little or no  
 CC effect.  
 SQ Sequence 25 AA;

Query Match 78.8%; Score 63; DB 1; Length 25;  
 Best Local Similarity 77.8%; Pred. No. 9.44e+00;  
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 13 YPADITLW 21  
 QY 1 YPAEITLYW 9  
 |||:|||||

RESULT 15  
 ID R71423 standard; peptide; 25 AA.  
 AC R71423;  
 DT 12-OCT-1995 (first entry)  
 DE Human MHC I alpha 3 domain peptide Dk-(197-221).  
 KW Major histocompatibility complex class I; MHC I; EGF receptor;  
 KW alpha 3 domain; peptide Dk-(197-221); interaction modulation;  
 KW arthritis; neoplasias; lupus erythematosus.  
 OS Homo sapiens.  
 PN WO9505189-A.  
 PD 23-FEB-1995.  
 PF 12-AUG-1994; U09189.  
 PR 12-AUG-1993; US-105416.  
 PA (REGC ) UNIV CALIFORNIA.  
 PI Goldstein A, Goodenow RS, Olsson L;  
 DR WPI; 95-098577/13.  
 PT Regulating cell surface receptor response - by modulating

PT Interaction between MHC class I antigen and the cell surface  
 PT receptor  
 PS Example 3; Page 38; 103pp; English.  
 CC R71420-R71423 are human major histocompatibility complex class 1  
 CC (MHC 1) derived peptides, they were used to modulate interactions  
 CC between MHC 1 and Egf cell surface receptors. Via competitive  
 CC inhibition the peptide diminishes the receptors response, this  
 CC feature may be useful for the treatment of neoplasias, lupus  
 CC erythematosus and arthritis.  
 SQ Sequence 25 AA;

Query Match 78.8%; Score 63; DB 1; Length 25;  
 Best Local Similarity 77.8%; Pred. No. 9.44e+00;  
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 13 YPADITLW 21  
 QY 1 YPAEITLYW 9  
 |||:|||||

Search completed: Fri Apr 14 23:18:41 2000  
 Job time : 41 secs.



PR 17-JUL-1997; US-896164.  
 PR 10-OCT-1997; US-061599.  
 PR 10-OCT-1997; US-061765.  
 PR 10-OCT-1997; US-048705.  
 PR 11-OCT-1997; GB-021697.  
 PA (LUDW-) LUDWIG INST CANCER RES.  
 PI (Chen Y, Gout I, Gure A, O'Hare M, Obata Y, Old LJ,  
 PI Pfeundschnuh M, Sahin U, Scanlan MJ, Stockert E,  
 PI Tureci O;  
 DR WPI: 99-132448/11.  
 PT New isolated cancer associated nucleic acids and polypeptides -  
 PT isolated using sera from cancer patients, used to develop products  
 PT for the diagnosis, monitoring or treatment of cancers  
 PS Disclosure: Page 417-418; 787pp; English.  
 CC The invention relates to a method for diagnosing a disorder characterised  
 CC by expression of a human cancer associated antigen precursor coded for by  
 CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a  
 CC biological sample isolated from a subject with an agent that specifically  
 CC binds to the NAM, an expression product or a fragment of an expression  
 CC product complexed with an HLA molecule; and (b) determining the  
 CC interaction between the agent and the NAM or the expression product as a  
 CC determination of the disorder. The products and methods can be used in  
 CC the diagnosis, monitoring, research, or treatment of conditions  
 CC characterised by the expression of various cancer associated antigens.  
 CC The invention provides nucleic acid sequences and encoded polypeptides  
 CC which are cancer associated antigen precursors expressed in human breast  
 CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and  
 CC lung cancer.  
 SQ Sequence 366 AA;

Query Match 80.0%; Score 64; DB 1; Length 366;  
 Best Local Similarity 88.9%; Pred. No. 7.35e+00;  
 Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 233 YPAEITLW 241  
 |||||  
 QY 1 YPAEITLW 9

RESULT 11  
 ID R12465 standard; Protein; 366 AA.  
 AC R12465;  
 DT 29-AUG-1991 (first entry)  
 DE HLA-C exon Cb-1.  
 KW Human leukocyte antigen; probe; major histocompatibility complex;  
 KW MHC; class I.  
 OS Homo sapiens.  
 PN J03112485-A.  
 PD 14-MAY-1991.  
 PF 22-SEP-1989; 247695.  
 PR 22-SEP-1989; JP-247695.  
 PA (OLYU) OLYMPUS OPTICAL KK.  
 DR WPI: 91-182989/25.  
 DR N-PSDB; Q12116.  
 PT HLA-C gene, DNA probe and transformant cells - for immunisation  
 PT of animals and monoclonal antibody development.  
 PS Claim 3; Page 2; 13pp; Japanese.  
 CC Probes comprising part of the DNA sequence encoding the protein can  
 CC be used to identify class I genes. The DNA can be expressed for  
 CC immunisation of animals and prodn. of monoclonal antibodies specific  
 CC for the HLA-C antigen. See also R12466 (same patent) and J03112486  
 CC and J03112487.  
 SQ Sequence 366 AA;

Query Match 80.0%; Score 64; DB 1; Length 366;  
 Best Local Similarity 88.9%; Pred. No. 7.35e+00;  
 Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 233 YPAEITLW 241  
 |||||  
 QY 1 YPAEITLW 9

RESULT 12  
 ID R12466 standard; Protein; 366 AA.  
 AC R12466;  
 DT 29-AUG-1991 (first entry)  
 DE HLA-C exon Cb-2.  
 KW Human leukocyte antigen; probe; major histocompatibility complex;  
 KW MHC; class I.  
 OS Homo sapiens.  
 PN J03112485-A.  
 PD 14-MAY-1991.  
 PF 22-SEP-1989; 247695.  
 PR 22-SEP-1989; JP-247695.  
 PA (OLYU) OLYMPUS OPTICAL KK.  
 DR WPI: 91-182989/25.  
 DR N-PSDB; Q12117.  
 PT HLA-C gene, DNA probe and transformant cells - for immunisation  
 PT of animals and monoclonal antibody development.  
 PS Claim 4; Page 2; 13pp; Japanese.  
 CC Probes comprising part of the DNA sequence encoding the protein can  
 CC be used to identify class I genes. The DNA can be expressed for  
 CC immunisation of animals and prodn. of monoclonal antibodies specific  
 CC for the HLA-C antigen. See also R12465 (same patent) and J03112486  
 CC and J03112487.  
 SQ Sequence 366 AA;

Query Match 80.0%; Score 64; DB 1; Length 366;  
 Best Local Similarity 88.9%; Pred. No. 7.35e+00;  
 Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 233 YPAEITLW 241  
 |||||  
 QY 1 YPAEITLW 9

RESULT 13  
 ID W68385 standard; Protein; 412 AA.  
 AC W68385;  
 DT 14-OCT-1998 (first entry)  
 DE Chimeric HLA-A2.1/beta-2 microglobulin protein.  
 KW Antigen; major histocompatibility complex; MHC; lymphocyte; detection;  
 KW immobilisation; cytotoxic T-cell; tumour; leukaemia; lymphoma;  
 KW viral infection; chimeric; beta-2 microglobulin.  
 OS Synthetic.  
 OS Homo sapiens.  
 PH Key  
 FT Domain 1, 302  
 FT Region 303, 312  
 FT Domain 313, 412  
 FT /note= "from HLA-A2.1 protein"  
 FT /note= "flexible linker"  
 FT /note= "from human beta-2 microglobulin"  
 PN W09744667-A2.  
 PD 27-NOV-1997.  
 PF 21-MAY-1997; F00892.  
 PR 21-MAY-1996; US-651925.  
 PA (INRM) INSERM INST NAT SANTE & RECH MEDICALE.  
 PA (INSP) INST PASTEUR.  
 PI Abastado J, Kourilsky P, Langlade-Demoyen P, Lone Y;  
 DR WPI: 98-018653/02.  
 DR N-PSDB; V30457.  
 PT Lymphocytes - for producing cytotoxic T cells for immuno-therapy of  
 PT cancers and viral infection  
 PS Example 2; Fig 10; 222pp; French.  
 CC Detection of antigen-specific lymphocytes comprises forming a complex  
 CC between antigenic peptides (see W68301-W68384 for examples) and  
 CC recombinantly produced major histocompatibility complex (MHC) molecules,  
 CC immobilising the complex and binding samples containing the  
 CC antigen-specific lymphocytes. Expression of the MHC molecule in a cell  
 CC is detected by antibodies against the MHC molecule or by an antibody to  
 CC a chimeric MHC-marker protein, e.g. an MHC/beta-2-microglobulin chimeric  
 CC protein. This sequence is an example of the chimeric protein and  
 CC comprises the first 3 domains of the HLA-A2.1 heavy chain linked to human

RESULT 6  
ID R03144 standard; protein; 362 AA.  
AC R03144;  
DT 19-MAR-1991 (first entry)  
DE Sequence of HLA-B\*51 antigen.  
KW Probe; HLA class I DNA; immunogen.  
OS Homo sapiens.  
PN EP-354580-A.  
PD 14-FEB-1990.  
PF 10-AUG-1989.  
PR 11-AUG-1988; JP-200758.  
PA (OLYU) Olympus Optical Co., Ltd.  
PI Kano K, Takiguchi;  
DR WPI: 90-046289/07.  
PT New DNA for class I human leucocyte antigens and derived probes and  
PT transformed cells, useful for DNA typing, as immunogens etc.  
PS Disclosure; Pages 12-13; 23pp; English.  
CC The HLA class I DNA can be used as a source of probes for use in DNA  
CC typing. Transformed cells, which are useful as immunogens, can be  
CC obtained by introducing these DNAs into eucaryotic cells.  
SQ Sequence 362 AA;  
Query Match 80.0%; Score 64; DB 1; Length 362;  
Best Local Similarity 88.9%; Pred. No. 7.35e+00;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Db 233 YPAEITLW 241  
| | | | | | | |  
QY 1 YPAEITLYW 9

RESULT 7  
ID R12463 standard; protein; 362 AA.  
AC R12463;  
DT 29-AUG-1991 (first entry)  
DE HLA-B\*53 exon.  
KW Human leucocyte antigen; probe; major histocompatibility complex;  
KW MHC; class I.  
OS Homo sapiens.  
PN J03112487-A.  
PD 14-MAY-1991.  
PR 22-SEP-1989; JP-247697.  
PA (OLYU) OLYMPUS OPTICAL KK.  
DR WPI: 91-182991/25.  
DR N-PSDB; Q12114.  
PT HLA-B\*53 gene, DNA probe and transformant cells - used for  
PT immunisation, identifying specificity of antiserum etc.  
PS Claim 2; Page 1; 11pp; Japanese.  
CC Probes comprising part of the sequence encoding the protein can be  
CC used to identify class I genes. The DNA can be expressed for  
CC immunisation of animals and prodn. of monoclonal antibodies  
CC specific for the HLA-B\*53 antigen. See also J03112485 and  
CC J03112486.  
SQ Sequence 362 AA;  
Query Match 80.0%; Score 64; DB 1; Length 362;  
Best Local Similarity 88.9%; Pred. No. 7.35e+00;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Db 233 YPAEITLW 241  
| | | | | | | |  
QY 1 YPAEITLYW 9

RESULT 8  
ID P70155 standard; protein; 362 AA.  
AC P70155;  
DT 10-MAR-1993 (revised)  
DT 03-APR-1991 (first entry)  
DE Sequence encoded by genomic DNA encoding human histocompatibility  
DE antigen HLA-B\*27.  
KW Ankylosing spondylitis; rheumatic disorder; diagnosis.

OS Homo sapiens.  
PN EP-226069-A.  
PD 24-JUN-1987.  
PF 21-NOV-1986; 116139.  
PR 01-JAN-1985; DE-542024.  
PR 21-DEC-1985; DE-545576.  
PA (BEHW) BEHRINGERWERKE AG.  
PI Szot's H. Weiss E, Dörner C, Lang M, Meo T, Riethmüller G;  
DR WPI: 87-171469/25.  
DR N-PSDB; N70225.  
PT DNA coding for human histocompatibility antigen HLA-B\*27 - useful  
PT for diagnosis and antigen and antibody prodn.  
PS Disclosure; p6; 13pp; German.  
CC The DNA may be used to detect the HLA-B\*27 gene (opt. mutated) in  
CC human genetic material. The HLA-B\*27 may be used to detect anti-HLA-  
CC B\*27 antibodies in human serum. The antibodies may be used to  
CC determine HLA-B\*27 levels in human serum, eg for diagnosis of  
CC rheumatic disorders, esp. ankylosing spondylitis.  
SQ Sequence 362 AA;  
Query Match 80.0%; Score 64; DB 1; Length 362;  
Best Local Similarity 88.9%; Pred. No. 7.35e+00;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Db 233 YPAEITLW 241  
| | | | | | | |  
QY 1 YPAEITLYW 9

RESULT 9  
ID R03142 standard; protein; 362 AA.  
AC R03142;  
DT 19-MAR-1991 (first entry)  
DE Sequence of HLA-B\*52 antigen.  
KW Probe; HLA class I DNA; immunogen.  
OS Homo sapiens.  
PN EP-354580-A.  
PD 14-FEB-1990.  
PF 10-AUG-1989.  
PR 11-AUG-1988; JP-200758.  
PA (OLYU) Olympus Optical Co., Ltd.  
PI Kano K, Takiguchi;  
DR WPI: 90-046289/07.  
PT New DNA for class I human leucocyte antigens and derived probes and  
PT transformed cells, useful for DNA typing, as immunogens etc.  
PS Disclosure; Page 13; 23pp; English.  
CC The HLA class I DNA can be used as a source of probes for use in DNA  
CC typing. Transformed cells, which are useful as immunogens, can be  
CC obtained by introducing these DNAs into eucaryotic cells.  
SQ Sequence 362 AA;  
Query Match 80.0%; Score 64; DB 1; Length 362;  
Best Local Similarity 88.9%; Pred. No. 7.35e+00;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Db 233 YPAEITLW 241  
| | | | | | | |  
QY 1 YPAEITLYW 9

RESULT 10  
ID Y07033 standard; protein; 366 AA.  
AC Y07033;  
DT 02-JUL-1999 (first entry)  
DE Breast cancer associated antigen precursor sequence.  
KW Cancer associated antigen; diagnosis; research; treatment; human;  
KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;  
KW prostate cancer.  
OS Homo sapiens.  
PN W09904265-A2.  
PD 28-JAN-1999.  
PF 15-JUL-1998; U14679.  
PR 22-JUN-1998; US-102322.

DE B53 self peptide cytotoxic T lymphocyte epitope.  
KW B53 self peptide; cytotoxic T; CTL; epitope; helper T; HTL; cell;  
KW lymphocyte; viruses; parasites; tumours; antigens; treatment;  
KW disease prevention.  
OS Homo sapiens.  
PN W09522317-A1.  
PD 24-AUG-1995.  
PF 16-FEB-1995; U02121.  
PR 16-FEB-1994; US-197484.  
PA (CYTE-) CYTEL CORP.  
PI Ceut RW Grey H. Sette AD, Vitello MA;  
DR WPI; 95-302545/39.  
PT Compn. inducing cytotoxic T lymphocyte response to pref. viral,  
PT bacterial, parasitic or tumour antigens - useful in the treatment  
PT and prevention of diseases associated with the antigen e.g.  
PT hepatitis B  
PS Disclosure; Page 17; 109pp; English.  
CC A compn. which induces a cytotoxic T lymphocyte (CTL) response to  
CC an antigen (Ag) in a mammal comprises, a CTL Ag response inducing  
CC peptide (i.e. R78824-R78853) and a lipid conjugated helper T cell  
CC inducing peptide. The compn. induces a CTL response to bacterial,  
CC viral or tumour Ags, and is therefore useful in the treatment and  
CC prevention of diseases associated with the Ag.  
SQ Sequence 9 AA;

Query Match 80.0%; Score 64; DB 1; Length 9;  
Best Local Similarity 88.9%; Pred. No. 7.35e+00;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 1 YPAEITLW 9  
| | | | | | | | | |  
QY 1 YPAEITLW 9

RESULT 3  
ID P80911 standard; protein; 274 AA.  
AC P80911;  
DT 18-SEP-1990 (first entry)  
DE Consensus sequence of peptides which constitute the alpha-1, alpha-2 and  
DE alpha-3 regions of a class I HLA molecule  
KW HLA-A2 epitopes; extracellular domains alpha-1, alpha-2 and alpha-3.  
OS Homo sapiens.  
FH Key Location/Qualifiers  
FT region /note="alpha-1 region"  
FT region 91..180  
FT region /note="alpha-2 region"  
FT region 181..272  
FT region /note="alpha-3 region"  
PN W08805784-A.  
PD 11-AUG-1988.  
PF 27-FEB-1988; U00245.  
PR 30-JAN-1987; US-008846.  
PA (STRD) Leland Stanford Jr Univ.  
PI Krensky AM, Parham P, Clayberger C;  
DR WPI; 88-235147/33.  
PT New peptide corresp. to major histocompatibility antigen regions -  
PT used for modulating cytotoxic T-lymphocyte activity in e.g.  
PT transplants or auto-immune diseases  
PS Example 9; Fig 4; 60pp; English.  
CC The consensus sequence is derived from a total of 23 HLA-A,B,C sequences.  
CC The protein sequences in the three extracellular domains (alpha-1,  
CC alpha-2 and alpha-3) are shown. The example concerned the effect of  
CC peptides from different HLA-A2 epitopes on cytolysis of target cells by  
CC CTL of different specificities.  
SQ Sequence 274 AA;

Query Match 80.0%; Score 64; DB 1; Length 274;  
Best Local Similarity 88.9%; Pred. No. 7.35e+00;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 209 YPAEITLW 217  
| | | | | | | | | |

QY 1 YPAEITLW 9

## RESULT 4

ID P70590 standard; protein; 337 AA.  
AC P70590;  
DT 10-APR-1991 (first entry)  
DE Sequence of the human histocompatibility antigen HLA B27.  
KW Rheumatic disorder; genetic screening; diagnosis;  
KW ankylosing spondylitis.  
OS Homo sapiens.  
PN DE3542024-A.  
PD 04-JUN-1987.  
PF 28-NOV-1985; 542024.  
PR 28-NOV-1985; DE-542024.  
PR 21-DEC-1985; DE-545576.  
PA (BEHW) BEHRINGER AG.  
PI Riethmuller G, Meo T, Weiss E, Szots H;  
DR WPI; 87-157893/23.  
DR N-PSDB; N70935.  
PT DNA coding for antigen HLA B27 - and diagnostic reagents contg.  
PT such DNA, antigen or antibody  
PS Disclosure; Page 5; 5pp; German.  
CC The DNA may be used as a hybridisation probe for detecting the HLA  
CC B27 gene, eg for assessing susceptibility to rheumatic disorders  
CC such as ankylosis spondylitis, or may be used to transform cells  
CC for prodn. of HLA B27. The HLA B27 may be used to detect HLA B27  
CC antibody in human serum, or to produce mono- or polyclonal HLA B27  
CC antibodies for use in immunoassay.  
SQ Sequence 337 AA;

Query Match 80.0%; Score 64; DB 1; Length 337;  
Best Local Similarity 88.9%; Pred. No. 7.35e+00;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 208 YPAEITLW 216  
| | | | | | | | | |  
QY 1 YPAEITLW 9

## RESULT 5

ID R12464 standard; Protein; 362 AA.  
AC R12464;  
DT 29-AUG-1991 (first entry)  
DE HLA-B35 antigen.  
KW Human leukocyte antigen; probe; major histocompatibility complex;  
KW MHC; class I.  
OS Homo sapiens.  
PN J03112486-A.  
PD 14-MAY-1991.  
PF 22-SEP-1989; 247697.  
PR 22-SEP-1989; JP-247697.  
PA (OLYU) OLYMPUS OPTICAL KK.  
DR WPI; 91-182991/25.  
DR N-PSDB; Q12115.  
PT HLA-B35 gene - used in DNA probe and transformant cells for  
PT immunising animals, for developing monoclonal antibody.  
PS Claim 1; Page 1; 11pp; Japanese.  
CC Probes comprising part of the sequence encoding this sequence can  
CC be used to identify Class I genes. The DNA can be expressed for  
CC immunisation of animals and prodn. of monoclonal antibodies  
CC specific for the HLA-B35 antigen. See also J03112485 and J03112487.  
SQ Sequence 362 AA;

Query Match 80.0%; Score 64; DB 1; Length 362;  
Best Local Similarity 88.9%; Pred. No. 7.35e+00;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 233 YPAEITLW 241  
| | | | | | | | | |  
QY 1 YPAEITLW 9

\*\*\*\*\*

W P S R L F (TM)

\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.

Copyright (c) 1993-1998 University of Edinburgh, U.K.

Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Apr 14 23:18:00 2000; Maspar time 7.54 Seconds

Tabular output not generated. 28.282 Million cell updates/sec

Title: &gt;US-08-452-843-4

Description: (1-9) from US08452843.pap

Perfect Score: 80

Sequence: 1 YPAEITLYW 9

Scoring table: PAM 150

Gap 15

Searched: 188963 seqs, 23686106 residues

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database: a-geneseq36

1:geneseqp

Statistics: Mean 18.465; Variance 56.813; scale 0.325

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	80	100.0	9	1 R89365	B53 self peptide deriv	1.17e-01
2	64	80.0	9	1 R78845	B53 self peptide cytot	7.35e-00
3	64	80.0	274	1 P80911	Consensus sequence of	7.35e-00
4	64	80.0	337	1 P70590	Sequence of the human	7.35e-00
5	64	80.0	362	1 R12464	HLA-B35 antigen.	7.35e-00
6	64	80.0	362	1 R03144	Sequence of HLA-B51 an	7.35e-00
7	64	80.0	362	1 R12463	HLA-Bw53 exon.	7.35e-00
8	64	80.0	362	1 P70135	Sequence encoded by ge	7.35e-00
9	64	80.0	362	1 R03142	Sequence of HLA-Bw52 a	7.35e-00
10	64	80.0	366	1 Y07033	Breast cancer associat	7.35e-00
11	64	80.0	366	1 R12465	HLA-C exon Cb-1.	7.35e-00
12	64	80.0	366	1 R12466	HLA-C exon Cb-2.	7.35e-00
13	64	80.0	412	1 W68385	Chimeric HLA-A2.1/beta	7.35e-00
14	63	78.8	25	1 R96622	MHC-I peptide Db-(197-	9.44e-00
15	63	78.8	25	1 R71423	Human MHC I alpha 3 do	9.44e-00
16	63	78.8	121	1 R52863	Mouse MHC alpha-3 doma	9.44e-00
17	58	72.5	547	1 W71526	Helicobacter polyepeti	3.25e-01
18	58	72.5	549	1 W55692	H. pylori ORF 06ge2050	3.25e-01
19	57	71.3	461	1 R75365	Phytase.	4.14e-01
20	56	70.0	131	1 W98363	H. pylori GHPO 1430 pr	5.28e-01
21	56	70.0	263	1 R80832	DM beta.	5.28e-01
22	55	68.8	547	1 W59481	Rat matrix metalloprot	6.71e-01
23	55	68.8	604	1 W10640	Membrane type matrix m	6.71e-01

24	55	68.8	607	1 W69480	Human matrix metallopr	6.71e+01
25	54	67.5	2466	1 R71498	Human protein tyrosine	8.53e+01
26	54	67.5	2466	1 W75999	Intracellular protein	8.53e+01
27	53	66.3	468	1 R25597	PHO.	1.08e+02
28	52	65.0	238	1 R93165	Anti-rhesus D recombin	1.37e+02
29	52	65.0	383	1 W60044	Human MHC class I chal	1.37e+02
30	52	65.0	385	1 W60043	Human MHC class I chal	1.37e+02
31	52	65.0	474	1 R14676	Rabbit vitronectin-lik	1.37e+02
32	52	65.0	760	1 W22213	Human transferrin rece	1.37e+02
33	51	63.8	667	1 R91240	B. cereus VIP1 protein	1.73e+02
34	51	63.8	667	1 R63794	Bacillus cereus 80 kDa	1.73e+02
35	51	63.8	667	1 W19510	B. cereus 80 kD VIP1A	1.73e+02
36	51	63.8	852	1 W19516	Maize optimised-B. cer	1.73e+02
37	51	63.8	852	1 R91246	VIP1A(a) protein with	1.73e+02
38	51	63.8	852	1 W46727	Maize optimised VIP1A	1.73e+02
39	51	63.8	880	1 W60224	Bacillus thuringiensis	1.73e+02
40	51	63.8	884	1 R91239	B. cereus VIP1A(a) ins	1.73e+02
41	51	63.8	1338	1 W19520	Maize optimised-B. cer	1.73e+02
42	51	63.8	1338	1 W46731	VIP2A(a)/VIP1A(a) fusi	1.73e+02
43	51	63.8	1338	1 R91247	VIP2A(a)-VIP1A(a) prot	1.73e+02
44	51	63.8	1346	1 W19513	B. cereus VIP1A(a)/VIP	1.73e+02
45	51	63.8	1346	1 R91245	VIP2A(a) and VIP1A(a)	1.73e+02

## ALIGNMENTS

## RESULT 1

ID R89365 standard; peptide; 9 AA.

AC R89365; (first entry)

DT 18-SEP-1996

DE B53 self peptide derived immunogenic peptide.

KW Immunogenic peptide; supermotif; HLA molecule; CTL response;

KW therapeutic; diagnostic; cancer; viral infection; hepatitis B;

KW hepatitis C.

OS Synthetic.

PN W09603140-A1.

PD 08-FEB-1996.

PF 21-JUL-1995; U09234.

PR 21-JUL-1994; US-278634.

PR 23-NOV-1994; US-344824.

PR 30-MAY-1995; US-452843.

PA (CYTE-) CYTEL CORP.

PI Sette A, Sidney J;

DR WPI: 96-116784/12.

PT Compn. comprising immunogenic peptide with supermotif allowing more than one HLA mol. to bind - used to induce CTL response in patient and for in vivo and ex vivo therapeutic and diagnostic applications

PS Claim 2; Page 26; 32pp; English.

CC The sequences given in R89362-82 are immunogenic peptides which were use in the composition of the invention. The composition comprises an immunogenic peptide of 9-10 residues with a supermotif which allows binding of more than one HLA molecule. It pref. comprises two conserved residues, a first at the 2nd position from the N-terminal is Pro, and a 2nd at the C-terminal is Met. These peptides are used to induce a CTL response in a patient. They are also useful in compositions for in vivo and ex vivo therapeutic and diagnostic applications, e.g. the treatment of cancer and viral infections, e.g. hepatitis B and C.

SQ Sequence 9 AA;

Query Match 100.0%; Score 80; DB 1; Length 9;

Best Local Similarity 100.0%; Pred. No. 1.17e-01;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 YPAEITLYW 9

QY 1 YPAEITLYW 9

RESULT 2

ID R78845 standard; peptide; 9 AA.

AC R78845;

DT 27-MAR-1996 (first entry)

THIS PAGE BLANK (USPTO)

RESULT	15
--------	----

DR PFAM; PF00109; ketoacyl-synt; 1.  
DR PFAM; PF00550; pp-binding; 1.  
KW Hypothetical protein; Fatty acid biosynthesis; Multifunctional enzyme;  
KW Phosphopantetheine; Transferase; Hydrolase; Oxidoreductase; Ligase;  
KW NADP; Membrane.  
FT DOMAIN ? ? ACYL TRANSFERASE.  
FT DOMAIN ? ? ENOYL REDUCTASE.  
FT DOMAIN ? ? BETA-KETOACYL REDUCTASE.  
FT DOMAIN ? 2100 ACYL CARRIER.  
FT DOMAIN 1 ? BETA-KETOACYL SYNTHASE.  
FT NP\_BIND 1567 1584 NADP (ER) (BY SIMILARITY).  
FT ACT\_SITE 1771 1786 NADP (KR) (BY SIMILARITY).  
FT ACT\_SITE 178 178 BETA-KETOACYL SYNTHASE (BY SIMILARITY).  
FT ACT\_SITE 624 624 ACYL TRANSFERASE (BY SIMILARITY).  
FT BINDING 2069 2069 PHOSPHOPANTHEINE (BY SIMILARITY).  
SQ SEQUENCE 2118 AA; 226495 MW; 7649D5A4 CRC32;

Query Match 55.28; Score 58; DB 2; Length 2118;  
Best Local Similarity 50.0%; Pred. No. 6.97e-01;  
Matches 5; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

Db 1099 LLDACFQSVI 1108  
QY 2 ILESCFRAVI 11  
I:|||||

RESULT 10  
ID O67106 PRELIMINARY; PRT; 145 AA.  
AC O67106;  
DT 01-AUG-1998 (TREMBLrel. 07, Created)  
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)  
DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)  
DE HYPOTHETICAL 16.6 KD PROTEIN.  
GN AQ\_978.  
OS Aquifex aeolicus.  
OC Bacteria; Aquificales; Aquificaceae; Aquifex.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=VF5;  
RX MEDLINE; 98196666.  
RA DECKERT G., WARREN P.V., GAASTERLAND T., YOUNG W.G., LENOX A.L.,  
RA GRAHAM D.E., OVERBECK R., SNEAD M.A., KELLER M., AUJAY M., HUBER R.,  
RA FELDMAN R.A., SHORT J.M., OLSON G.J., SWANSON R.V.;  
RT "The complete genome of the hyperthermophilic bacterium Aquifex  
aeolicus.";  
RL Nature 392:353-358(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=VF5;  
RA DECKERT G., WARREN P.V., GAASTERLAND T., YOUNG W.G., LENOX A.L.,  
RA GRAHAM D.E., OVERBECK R., SNEAD M.A., KELLER M., AUJAY M., HUBER R.,  
RA FELDMAN R.A., SHORT J.M., OLSON G.J., SWANSON R.V.;  
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AE000716; AAC07068.1;  
KW Hypothetical protein.  
SQ SEQUENCE 145 AA; 16638 MW; 83C4267C CRC32;

Query Match 64.08; Score 57; DB 2; Length 145;  
Best Local Similarity 62.58; Pred. No. 1.13e+00;  
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 9 CLLECCYR 16  
QY 1 CILESCFR 8  
I:|||||

RESULT 11  
ID P89938 PRELIMINARY; PRT; 1727 AA.  
AC P89938;  
DT 01-MAY-1997 (TREMBLrel. 03, Created)  
DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)  
DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)  
DE REPLICASE ORF1A POLYPROTEIN.

OS Equine arteritis virus (EAV).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;  
OC Arteriviridae; Arterivirus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 97175715.  
RA VAN DINTEN L.C., DEN BOON J.A., WASSENAR A.L.M., SPAAN W.J.M.,  
RA SNIJDER E.J.;  
RT "An infectious arterivirus cDNA clone: Identification of a replicase  
point mutation that abolishes discontinuous mRNA transcription.";  
RL Proc. Natl. Acad. Sci. U.S.A. 94:991-996(1997).  
DR EMBL; Y07862; CAA69186.1;  
KW Polyprotein.  
SQ SEQUENCE 1727 AA; 186985 MW; A602D83D CRC32;

Query Match 64.08; Score 57; DB 14; Length 1727;  
Best Local Similarity 60.0%; Pred. No. 1.13e+00;  
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 344 CLDESCFRGI 353  
QY 1 CILESCFRAV 10  
I:|||||

RESULT 12  
ID P89939 PRELIMINARY; PRT; 3175 AA.  
AC P89939;  
DT 01-MAY-1997 (TREMBLrel. 03, Created)  
DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)  
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)  
DE REPLICASE ORF1B POLYPROTEIN.  
OS Equine arteritis virus (EAV).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;  
OC Arteriviridae; Arterivirus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 97175715.  
RA VAN DINTEN L.C., DEN BOON J.A., WASSENAR A.L.M., SPAAN W.J.M.,  
RA SNIJDER E.J.;  
RT "An infectious arterivirus cDNA clone: Identification of a replicase  
point mutation that abolishes discontinuous mRNA transcription.";  
RL Proc. Natl. Acad. Sci. U.S.A. 94:991-996(1997).  
DR EMBL; Y07862; CAA69187.1;  
KW Polyprotein.  
SQ SEQUENCE 3175 AA; 345376 MW; FDFD6351 CRC32;

Query Match 64.08; Score 57; DB 14; Length 3175;  
Best Local Similarity 60.0%; Pred. No. 1.13e+00;  
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 344 CLDESCFRGI 353  
QY 1 CILESCFRAV 10  
I:|||||

RESULT 13  
ID Q9XIB4 PRELIMINARY; PRT; 210 AA.  
AC Q9XIB4;  
DT 01-NOV-1999 (TREMBLrel. 12, Created)  
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)  
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)  
DE F13F21.9 PROTEIN.  
GN F13F21.9.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC eudicotyledons; Spermatophyta; Magnoliophyta; eudicotyledons;  
OC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;  
OC Arabidopsis.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA PEDERSPIEL N.A., PALM C.J., CONWAY A.B., CONN L., HANSEN N.F.,  
RA ALTAFI H., ARAUJO R., HUIZAR L., ROWLEY D., BUEHLER E., DUNN P.,  
RA GONZALEZ A., KRENETSKAIA I., KIM C., LENZ C., LI J., LIU S.,

```
DR EMBL; AF105220; AAD45228.1; -.
SQ SEQUENCE 622 AA; 70378 MW; 76A06AA5 CRC32;

Query Match      65.2%; Score 58; DB 14; Length 622;
Best Local Similarity 60.0%; Pred. No. 6.97e-01;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 330 CILTNCRGV 339
|||||:|:|:|
QY 1 CILESCFRAV 10

RESULT 6
ID Q9YNAO PRELIMINARY; PRT; 622 AA.
AC Q9YNAO;
DT 01-MAY-1999 (TReMBLrel. 10, Created)
DT 01-MAY-1999 (TReMBLrel. 10, Last sequence update)
DT 01-MAY-1999 (TReMBLrel. 10, Last annotation update)
DE ENV PROTEIN.
GN ENV.
OS Sheep pulmonary adenomatosis virus (Jaagsiekte sheep retrovirus)
OS (JSRV).
OS Viruses; Retrov. viruses; Retroviridae; Type D retroviruses.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=809T;
RA BAI J., HARRISON J.V., CARLSON J.O., DEMARTINI J.C.;
RT "Jaagsiekte sheep retrovirus isolates from ovine pulmonary carcinomas
RT exhibit two envelope genotypes and a conserved x gene of unknown
RT function.";
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; Y18305; CAA71121.1; -.
SQ SEQUENCE 622 AA; 70418 MW; 7A04BFDB CRC32;

Query Match      65.2%; Score 58; DB 14; Length 622;
Best Local Similarity 60.0%; Pred. No. 6.97e-01;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 330 CILTNCRGV 339
|||||:|:|:|
QY 1 CILESCFRAV 10

RESULT 7
ID Q9WJR2 PRELIMINARY; PRT; 622 AA.
AC Q9WJR2;
DT 01-NOV-1999 (TReMBLrel. 12, Created)
DT 01-NOV-1999 (TReMBLrel. 12, Last sequence update)
DT 01-NOV-1999 (TReMBLrel. 12, Last annotation update)
DE ENV PROTEIN.
GN ENV.
OS Sheep pulmonary adenomatosis virus (Jaagsiekte sheep retrovirus).
OS Viruses; Retrov. viruses; Retroviridae; Type D retroviruses.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JS7;
RA BAI J., HARRISON J.V., CARLSON J.O., DEMARTINI J.C.;
RT "Jaagsiekte sheep retrovirus isolates from ovine pulmonary carcinomas
RT exhibit two envelope genotypes and a conserved x gene of unknown
RT function.";
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; Y18305; CAA71121.1; -.
SQ SEQUENCE 622 AA; 70336 MW; 7C6A3E99 CRC32;

Query Match      65.2%; Score 58; DB 14; Length 622;
Best Local Similarity 60.0%; Pred. No. 6.97e-01;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 330 CILTNCRGV 339
|||||:|:|:|
QY 1 CILESCFRAV 10

RESULT 8
ID Q9XW22 PRELIMINARY; PRT; 936 AA.
AC Q9XW22;
DT 01-NOV-1999 (TReMBLrel. 12, Created)
DT 01-NOV-1999 (TReMBLrel. 12, Last sequence update)
DT 01-NOV-1999 (TReMBLrel. 12, Last annotation update)
DE Y18D10A.7 PROTEIN.
GN Y18D10A.7.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
OC Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 94150718.
RA HARRIS B.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SWALDON N., SMITH A., SONHAMMER E., STADEN R., SULSTON J.,
RA THERRY-MIEG J., THOMAS K., VAUGHAN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
DR EMBL; AL034393; CAA22312.1; -.
SQ SEQUENCE 936 AA; 106623 MW; 39A7FF0F CRC32;

Query Match      65.2%; Score 58; DB 5; Length 936;
Best Local Similarity 40.0%; Pred. No. 6.97e-01;
Matches 4; Conservative 6; Mismatches 0; Indels 0; Gaps 0;

Db 889 VVQTCFRTVV 898
|||||:|:|:|
QY 2 ILESCFRAVI 11

RESULT 9
ID Q49624 PRELIMINARY; PRT; 2118 AA.
AC Q49624;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-MAY-1999 (TReMBLrel. 10, Last annotation update)
DE PROBABLE MYCOGEROSIC ACID SYNTHASE.
GN MASA OR B1170_C2_209.
OS Mycobacterium leprae.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
RN [1]
RP SEQUENCE FROM N.A.
RA ROBISON K., SMITH D.R.;
RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: CATALYZES THE ELONGATION OF N-FATTY ACYL-COA WITH
CC METHYLMALONYL-COA (NOT MALONYL-COA) AS THE ELONGATING AGENT TO
CC FORM MYCOGEROSYL LIPIDS
CC COFACTOR: CONTAINS ONE COVALENTLY BOUND PHOSPHOPANTHETHEINE.
CC -1- SUBUNIT: HOMODIMER WHOSE MONOMERS PROBABLY HAVE A HEAD TO TAIL
CC ARRANGEMENT.
CC -1- SUBCELLULAR LOCATION: MEMBRANE-ASSOCIATED.
CC -1- SIMILARITY: PARTIAL TO S. ERYTHRAEA ERYTHRONOLIDE SYNTHASE, MODULE
CC 4, AND TO VERTEBRATE FATTY ACID SYNTHASES.
DR EMBL; U00010; AAA17069.1; -.
DR PFAM; PF00698; Acyl_transf; 1.
DR PFAM; PF00107; adh_zinc; 1.
```



KW Receptor; Transcription regulation; DNA-binding; Nuclear protein;  
Zinc-finger.  
SQ SEQUENCE 232 AA; 26943 MW; 84973D12 CRC32;  
Query Match 67.4%; Score 60; DB 5; Length 232;  
Best Local Similarity 77.8%; Pred. No. 2.61e-01;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 72 CRLEKCFRA 80  
| | | | |  
QY 1 CILESCFRA 9

RESULT 2  
ID Q9YDS5 PRELIMINARY; PRT; 149 AA.  
AC Q9YDS5;  
DT 01-NOV-1999 (TREMBlrel. 12, Created)  
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)  
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)  
DE 149AA LONG HYPOTHETICAL PROTEIN.  
GN APE0842.  
OS Aeropyrum pernix.  
OC Archaea; Crenarchaeota; Aeropyrum.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-K1;  
RX MEDLINE; 99310339.  
RA KAWARABAYASI Y., HINO Y., HORIKAWA H., YAMAZAKI S., HAIKAWA Y.,  
RA JIN-NO K., TAKAHASHI M., SEKINE M., BABA S., ANKAI A., KOSUGI H.,  
RA HOSOYAMA A., FUKUI S., NAGAI Y., NISHIJIMA K., NAKAZAWA H.,  
RA TAKAMIYA M., MASUDA S., FUNAHASHI T., TANAKA T., KUDOH Y.,  
RA YAMAZAKI J., KUSHIDA N., OGUCHI A., AOKI K., KUBOTA K., NAKAMURA Y.,  
RA NOMURA N., SAKO Y., KIKUCHI H.;  
RT "Complete genome sequence of an aerobic hyper-thermophilic  
RT crenarchaeon, Aeropyrum pernix K1.";  
RL DNA Res. 6:83-101(1999).  
DR EMBL; AP000060; BAA79822.1; -;  
SQ SEQUENCE 149 AA; 15892 MW; 50930B2C CRC32;

Query Match 65.2%; Score 58; DB 1; Length 149;  
Best Local Similarity 55.6%; Pred. No. 6.97e-01;  
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 86 CILAACFWS 94  
| | | | |  
QY 1 CILESCFRA 9

RESULT 3  
ID Q9WJ21 PRELIMINARY; PRT; 617 AA.  
AC Q9WJ21;  
DT 01-NOV-1999 (TREMBlrel. 12, Created)  
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)  
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)  
DE ENV PROTEIN.  
GN ENV.  
OS Ovine enzootic nasal tumour virus.  
OC Viruses; Retrovirdae; Retroviridae.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-SHEEP TNO28;  
RX MEDLINE; 99214337.  
RA COUSINS C.A.M., MINGUIJON E., DALZIEL R.G., ORTIN A., GARCIA M.,  
RA PARK J., GONZALEZ L., SHARP J.M., DE LAS HERAS M.;  
RT "Complete sequence of enzootic nasal tumor virus, a retrovirus  
RT associated with transmissible intranasal tumors of sheep.";  
RL J. Virol. 73:3986-3993(1999).  
DR EMBL; Y16627; CAB41418.1; -;  
FT CHAIN 1 378 ENV SU PROTEIN.  
FT CHAIN 379 617 ENV TM PROTEIN.  
SQ SEQUENCE 617 AA; 69710 MW; C98AE6EE CRC32;

Query Match 65.2%; Score 58; DB 14; Length 617;

Best Local Similarity 60.0%; Pred. No. 6.97e-01;  
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 323 CILTNCIRGV 332  
| | | | |  
QY 1 CILESCFRA 10

RESULT 4  
ID Q9YN97 PRELIMINARY; PRT; 617 AA.  
AC Q9YN97;  
DT 01-MAY-1999 (TREMBlrel. 10, Created)  
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)  
DT 01-MAY-1999 (TREMBlrel. 10, Last annotation update)  
DE ENV PROTEIN.  
GN ENV.  
OS Sheep pulmonary adenomatosis virus (Jaagsiekte sheep retrovirus)  
OC Viruses; Retrovirdae; Retroviridae; Type D retroviruses.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-84RS28;  
RA BAI J., HARRISON J.V., CARLSON J.O., DEMARTINI J.C.;  
RT "Jaagsiekte sheep retrovirus isolates from ovine pulmonary carcinomas  
RT exhibit two envelope genotypes and a conserved X gene of unknown  
RT function.";  
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL; Y18304; CAA77120.1; -;  
SQ SEQUENCE 617 AA; 69535 MW; B7594648 CRC32;

Query Match 65.2%; Score 58; DB 14; Length 617;  
Best Local Similarity 60.0%; Pred. No. 6.97e-01;  
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 323 CILTNCIRGV 332  
| | | | |  
QY 1 CILESCFRA 10

RESULT 5  
ID Q9YJ35 PRELIMINARY; PRT; 622 AA.  
AC Q9YJ35;  
DT 01-MAY-1999 (TREMBlrel. 10, Created)  
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)  
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)  
DE ENV PROTEIN.  
GN ENV.  
OS Sheep pulmonary adenomatosis virus (Jaagsiekte sheep retrovirus)  
OC Viruses; Retrovirdae; Retroviridae; Type D retroviruses.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-83RS28, JS7;  
RA BAI J., HARRISON J.V., CARLSON J.O., DEMARTINI J.C.;  
RT "Jaagsiekte sheep retrovirus isolates from ovine pulmonary carcinomas  
RT exhibit two envelope genotypes and a conserved X gene of unknown  
RT function.";  
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL; Y18301; CAA77115.1; -;  
SQ SEQUENCE FROM N.A.  
RC STRAIN-JSRV21;  
RX MEDLINE; 99329222.  
RA PALMARINI M., SHARP J.M., DE LAS HERAS M., FAN H.;  
RT "Jaagsiekte sheep retrovirus is necessary and sufficient to induce a  
RT contagious lung cancer in sheep.";  
RL J. Virol. 73:6964-6972(1999).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-JSRV21;  
RA PALMARINI M., SHARP J.M., FAN H.;  
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL; Y18303; CAA77119.1; -;  
DR EMBL; Y18301; CAA77115.1; -;



THIS PAGE BLANK (USPTO)

RA WALSH S.V., WHITEHEAD S.;  
RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; Z38059; CAA86137.1; -  
DR PIR; S48393; S48393.  
KW Hypothetical protein.  
SQ SEQUENCE 129 AA; 14437 MW; 48F132D8 CRC32;

Query Match 58.4%; Score 52; DB 1; Length 129;  
Best Local Similarity 50.0%; Pred. No. 5.92e+00;  
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 7 VLEPCKNVI 16  
:||:|: ||  
Qy 2 ILESCFRAVI 11

Search completed: Fri Apr 14 23:14:03 2000  
Job time : 41 secs.

Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 357 CTGECFKA 365  
| :||:|  
Qy 1 CILESCFA 9

```
RESULT 13
ID CID_DROME STANDARD; PRT; 1377 AA.
AC P19538;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE CUBITUS INTERRUPTUS DOMINANT PROTEIN.
GS CI-D.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-OREGON-R;
RX MEDLINE; 90346286.
RA ORENIC T.V., SLOJARSKI D.C., KROLL K.L., HOLMGREN R.A.;
RT "Cloning and characterization of the segment polarity gene cubitus
interruptus dominant of Drosophila.";
RL Genes Dev. 4:1053-1067(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-OREGON-R;
RX MEDLINE; 92146935.
RA BERRY A.J., ADIORA J.W., KREITMAN M.;
RT "Lack of polymorphism on the Drosophila fourth chromosome resulting
from selection.";
RL Genetics 129:1111-1117(1991).
CC -!- FUNCTION: INVOLVED IN SEGMENT POLARITY. IS REQUIRED FOR THE NORMAL
DEVELOPMENT OF THE POSTERIOR HALF OF EACH EMBRYONIC SEGMENT.
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- DEVELOPMENTAL STAGE: EXPRESSED UNIFORMLY THROUGHOUT THE BLASTODERM
STAGE AND GASTRULATION AND DOES NOT RESOLVE INTO SEGMENTALLY
REPEATING STRIPES UNTIL THE END OF THE SHORT PHASE OF GERM-BAND
EXTENSION.
CC -!- SIMILARITY: TO THE GLI-RELATED GROUP OF C2H2-TYPE ZINC-FINGERS
PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
CC EMBL; X54360; CAA38244.1; -.
DR PIR; A35817; A35817.
DR HSSP; P08151; 2GLI.
DR FLYBASE; FBgn004859; ci.
DR PROSITE; PS00028; ZINC_FINGER_C2H2; 4.
DR PFAM; PF00096; zf-C2H2; 5.
KW Developmental protein; Segmentation polarity protein; Zinc-finger;
Metal-binding; DNA-binding; Repeat; Nuclear protein.
FT DOMAIN 451 603 ZINC-FINGERS.
FT ZN_FING 451 476 C2H2-TYPE.
FT ZN_FING 484 511 C2H2-TYPE.
FT ZN_FING 517 541 C2H2-TYPE.
FT ZN_FING 547 572 C2H2-TYPE.
FT ZN_FING 578 603 C2H2-TYPE.
SQ SEQUENCE 1377 AA; 150881 MW; A14EB3FC CRC32;
```

Query Match 59.6%; Score 53; DB 1; Length 1377;  
Best Local Similarity 55.8%; Pred. No. 3.76e+00;  
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 519 CTGECFKA 527  
| :||:|  
Qy 1 CILESCFA 9

```
RESULT 14
ID VA4_SOLIN STANDARD; PRT; 117 AA.
AC P35777;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE VENOM ALLERGEN IV (ALLERGEN SOL I 4) (SOL I IV).
OS Solenopsis invicta (Red imported fire ant).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata;
OC Formicoidea; Formicidae; Solenopsis.
RN [1]
RP SEQUENCE.
RC TISSUE-VENOM;
RX MEDLINE; 93139387.
RA HOFFMAN D.R.;
RT "Allergens in Hymenoptera venom XXIV: the amino acid sequences of
imported fire ant venom allergens Sol I II, Sol I III, and Sol I
IV.";
RL J. Allergy Clin. Immunol. 91:71-78(1993).
RN [2]
RP PARTIAL SEQUENCE OF 1-31.
RC TISSUE-VENOM;
RX MEDLINE; 90285439.
RA HOFFMAN D.R., SMITH A.M., SCHMIDT M., MOFFITT J.E., GURALNICK M.;
RT "Allergens in Hymenoptera venom. XXII. Comparison of venoms from two
species of imported fire ants, Solenopsis invicta and richteri.";
RL J. Allergy Clin. Immunol. 85:988-996(1990).
CC -!- DISEASE: THE MOST COMMON CAUSE OF INSECT VENOM ALLERGY IN THE
SOUTHEASTERN UNITED STATES IS THE IMPORTED FIRE ANT.
CC -!- SIMILARITY: MONOMER.
CC -!- SIMILARITY: TO VENOM ALLERGEN II.
DR PIR; C37330; C37330.
DR PIR; B44582; B44582.
KW Venom; Allergen.
FT VARIANT 23 23 R -> H.
FT VARIANT 37 37 L -> I.
SQ SEQUENCE 117 AA; 13341 MW; 7174AB01 CRC32;
```

Query Match 58.4%; Score 52; DB 1; Length 117;  
Best Local Similarity 66.7%; Pred. No. 5.92e+00;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 12 ILEKIRTV 20  
| :||:|  
Qy 2 ILESCFAV 10

```
RESULT 15
ID Y101_YEAST STANDARD; PRT; 129 AA.
AC P40461;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE HYPOTHETICAL 14.4 KD PROTEIN IN CCT2-AML2 INTERGENIC REGION.
GN Y11141W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
OC Saccharomycetaceae; Saccharomycetes.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-S288C / AB972;
RA BARRELL B.G., BADCOCK K., BANKIER A.T., BOWMAN S., BROWN D.,
RA CHURCHER C.M., CONNOR R., COPSEY T., DEAR S., DEVLIN K., FRASER A.,
RA GENTLES S., HAMLYN N., HORSNELL T.S., HUNT S., JAGELS K., JONES M.,
RA LOUIS E., LYE G., MOULE S., MOULE T., ODELL C., PEARSON D.,
RA RAJANDREAM M.A., RILES L., ROWLEY N., SKELTON J., SMITH V.,
```

FT CONFLICT 431 431 G -> A (IN REF. 1).  
FT CONFLICT 436 436 P -> S (IN REF. 2).  
SQ SEQUENCE 465 AA; 51740 MW; D3008841 CRC32;

Query Match 59.6%; Score 53; DB 1; Length 465;  
Best Local Similarity 66.7%; Pred. No. 3.76e+00;  
Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Db 106 CRLKKCFRA 114  
| | | | |  
QY 1 CILESCFRA 9

RESULT 11  
ID NUC2\_SCHPO STANDARD; PRT; 665 AA.  
AC P10505;  
DT 01-JUL-1989 (Rel. 11, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 01-OCT-1996 (Rel. 34, Last annotation update)  
DE NUCLEAR SCAFFOLD-LIKE PROTEIN P76.  
GN NUC2 OR SPAC17C9.01C.  
OS Schizosaccharomyces pombe (Fission yeast).  
OC Eukaryota; Fungi; Ascomycota; Archiascomycetes;  
OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
OC Schizosaccharomycetes.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-TS MUTANT NUC2-663;  
RX MEDLINE; 88198361.

RA HIRANO M., HIRAKA Y., YANAGIDA M.;  
RT "A temperature-sensitive mutation of the Schizosaccharomycetes pombe  
gene nuc2+ that encodes a nuclear scaffold-like protein blocks  
spindle elongation in mitotic anaphase.";  
RL J. Cell Biol. 106:1171-1183(1988).  
RN [2]  
RP REVISION TO 649.  
RA YANAGIDA M.;  
RL Submitted (MAR-1989), to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE OF 1-557 FROM N.A.

RC STRAIN-972;  
RA MURPHY L., MCDUGALL R., JONES L., SIMPSON I., MCNEIL A., HARRIS D.,  
RA BARRELL B.G., RAJANDRAM M.A., WALSH S.V.;  
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP DOMAINS.  
RX MEDLINE; 90124640.

RA HIRANO M., KINOSHITA N., MORIKAWA K., YANAGIDA M.;  
RT "Snap helix with knob and hole: essential repeats in S. pombe nuclear  
protein nuc2+.";  
RL Cell 60:319-328(1990).  
CC -1- FUNCTION: NUC2 INTERACTS WITH SPINDLE APPARATUS, CHROMOSOMES,  
OR NUCLEAR ENVELOPE, AND INTERCONNECT NUCLEAR AND CYTOSKELETAL  
FUNCTIONS IN MITOSIS, SO THE ELONGATION OF THE SPINDLE IN ANAPHASE  
IS BLOCKED.  
CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
CC -1- SIMILARITY: CONTAINS 10 TPR DOMAINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; X07693; CAA30532.1; -.  
DR EMBL; Z73099; CAA97347.1; -.  
DR PIR; A30185; A30185.

DR PFAM; PF00515; TPR; 7.  
KW Cell division; Cell cycle; Mitosis; Repeat; TPR domain;  
Nuclear protein.  
FT REPEAT 118 151 TPR 1.

FT DNA\_BIND 191 257  
FT REPEAT 332 365 TPR 2.  
FT REPEAT 366 399 TPR 3.  
FT REPEAT 400 433 TPR 4.  
FT REPEAT 434 467 TPR 5.  
FT REPEAT 468 501 TPR 6.  
FT REPEAT 502 535 TPR 7.  
FT REPEAT 536 569 TPR 8.  
FT REPEAT 570 603 TPR 9.  
FT REPEAT 604 637 TPR 10.  
FT MUTAGEN 504 504 G->D: IN TEMPERATURE SENSITIVE MUTANT.  
FT CONFLICT 440 440 C -> W (IN REF. 1).  
SQ SEQUENCE 665 AA; 76171 MW; D46BC3C8 CRC32;

Query Match 59.6%; Score 53; DB 1; Length 665;  
Best Local Similarity 71.4%; Pred. No. 3.76e+00;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 435 CILANCF 441  
| | | | |  
QY 1 CILESCF 7

RESULT 12  
ID GLI4\_XENLA STANDARD; PRT; 1361 AA.  
AC Q91661;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 15-JUL-1998 (Rel. 36, Last annotation update)  
DE ZINC FINGER PROTEIN GLI4 (NEURAL SPECIFIC DNA BINDING PROTEIN XGLI4)  
DE (XGLI-4).  
GN GLI4.

OS Xenopus laevis (African clawed frog).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Amphibia;  
OC Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae; Xenopodinae;  
OC Xenopus.  
RN [1]  
RP SEQUENCE FROM N.A.

RX MEDLINE; 97346726.  
RA MARINE J.C., BELLEFROID E.J., PENDEVILLE H., MARTIAL J.A., PIELER T.;  
RT "A role for xenopus Gli-type zinc finger proteins in the early  
embryonic patterning of mesoderm and neuroectoderm.";  
RL Mech. Dev. 63:211-225(1997).  
CC -1- FUNCTION: HAS AN ESSENTIAL ROLE IN THE EARLY EMBRYONIC PATTERNING  
OF MESODERM AND NEUROECTODERM.

CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).  
CC -1- SIMILARITY: TO THE GLI-RELATED GROUP OF C2H2-TYPE ZINC-FINGERS  
PROTEINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; U42462; AAA98467.1; -.  
DR HSSP; P08151; ZGLI.  
DR PROSITE; P500028; ZINC\_FINGER\_C2H2; 4.  
DR PFAM; PF00096; zf-C2H2; 5.  
KW Zinc-finger; Metal-binding; DNA-binding; Transcription regulation;

KW Nuclear protein.  
FT DOMAIN 289 441 ZINC-FINGERS.  
FT ZN\_FING 289 314 C2H2-TYPE.  
FT ZN\_FING 322 349 C2H2-TYPE.  
FT ZN\_FING 355 379 C2H2-TYPE.  
FT ZN\_FING 385 410 C2H2-TYPE.  
FT ZN\_FING 416 441 C2H2-TYPE.  
SQ SEQUENCE 1361 AA; 149554 MW; 70E6495C CRC32;

Query Match 59.6%; Score 53; DB 1; Length 1361;  
Best Local Similarity 55.6%; Pred. No. 3.76e+00;

15-JUL-1999 (Rel. 38, Last annotation update)  
DE HEPATOCYTE NUCLEAR FACTOR 4-ALPHA (HNF-4-ALPHA) (TRANSCRIPTION FACTOR  
DE HNF-4) (TRANSCRIPTION FACTOR 14).  
GN HNF4A OR NR2A1 OR TC1F4 OR HNF4.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Primates; Catarrhini; Homiidae; Homo.  
RN [1]  
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.  
RC TISSUE=LIVER;  
RX MEDLINE: 95011627.  
RA CHARTIER F.B., BOSSU J.-P., LAUDET V., FRUCHART J.-C., LAINE B.;  
RT "Cloning and sequencing of cDNAs encoding the human hepatocyte  
RT nuclear factor 4 indicate the presence of two isoforms in human  
RT liver.";  
RL Gene 147:269-272(1994).  
RN [2]  
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.  
RC TISSUE=LIVER;  
RX MEDLINE: 97082982.  
RA KRITIS A.A., ARGYROKASTRITIS A., MOSCHONAS N.K., POWER S.,  
RA KATRAKILI N., ZANNIS V.I., CERECHINI S., TALIANIDIS I.;  
RT "Isolation and characterization of a third isoform of human hepatocyte  
RT nuclear factor 4.";  
RL Gene 173:275-280(1996).  
RN [3]  
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.  
RC TISSUE=KIDNEY;  
RX MEDLINE: 96182096.  
RA DREWS T., SENKEL S., HOLEWA B., RYFFEL G.U.;  
RT "Human hepatocyte nuclear factor 4 isoforms are encoded by distinct  
RT and differentially expressed genes.";  
RL Mol. Cell. Biol. 16:925-931(1996).  
RN [4]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 97100944.  
RA YAMAGATA K., FURUTA H., ODA N., KAISAKI P.J., MENZEL S., COX N.J.,  
RA FAJANS S.S., SIGNORINI S., STOFFEL M., BELL G.I.;  
RT "Mutations in the hepatocyte nuclear factor-4alpha gene in maturity-  
RT onset diabetes of the young (MODY1).";  
RL Nature 384:458-460(1996).  
RN [5]  
RP VARIANT MODY1 TRP-127.  
RX MEDLINE: 97458990.  
RA FURUTA H., IWASAKI N., ODA N., HINOKIO Y., HORIKAWA Y., YAMAGATA K.,  
RA YANO N., SUGAHRO J., OGATA M., OHGAWARA H., OMORI Y., IWAMOTO Y.,  
RA BELL G.I.;  
RT "Organization and partial sequence of the hepatocyte nuclear factor-4-  
RT alpha/MODY1 gene and identification of a missense mutation, R127W, in  
RT a Japanese family with MODY.";  
RL Diabetes 46:1652-1657(1997).  
RN [6]  
RP VARIANT MODY1 ILE-393.  
RX MEDLINE: 98119841.  
RA HANI E.H., SUAUD L., BOUTIN P., CHEVRE J.-C., DURAND E., PHILIPPI A.,  
RA DEMENAIS F., VIONNET N., FURUTA H., VELHO G., BELL G.I., LAINE B.,  
RA FROGUEL P.;  
RT "A missense mutation in hepatocyte nuclear factor-4-alpha, resulting  
RT in a reduced transactivation activity, in human late-onset non-  
RT insulin-dependent diabetes mellitus.";  
RL J. Clin. Invest. 101:521-526(1998).  
CC -1- FUNCTION: TRANSCRIPTIONALLY CONTROLLED TRANSCRIPTION FACTOR. BINDS  
CC TO DNA SITES REQUIRED FOR THE TRANSCRIPTION OF ALPHA 1-  
CC ANTITRYPSIN, APOLOPROTEIN CIII, TRANSFERRIN GENES AND HNF1-  
CC ALPHA. MAY BE ESSENTIAL FOR DEVELOPMENT OF THE LIVER, KIDNEY AND  
CC INTESTINE.  
CC -1- SUBUNIT: HOMODIMERIZATION IS REQUIRED FOR HNF4-ALPHA TO BIND TO  
CC ITS RECOGNITION SITE.  
CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
CC -1- ALTERNATIVE PRODUCTS: AT LEAST FOUR ISOFORMS: HNF4-ALPHA-1/HNF-4B  
CC (SHOWN HERE), HNF4-ALPHA-2/HNF4-A, HNF4-ALPHA-3/HNF4-C AND HNF4-  
CC ALPHA-4; ARE PRODUCED BY ALTERNATIVE SPLICING.  
CC -1- DISEASE: DEFECTS IN TC1F4 ARE A CAUSE OF MONOGENIC AUTOSOMAL

CC DOMINANT NON-INSULIN-DEPENDENT DIABETES MELLITUS TYPE I (MODY1  
CC OR MODY-1) (ALSO KNOWN AS NIDDM). A FORM OF DIABETES THAT IS  
CC CHARACTERIZED BY AN AUTOSOMAL DOMINANT MODE OF INHERITANCE, ONSET  
CC DURING CHILDHOOD (USUALLY BEFORE 25 YEARS OF AGE) AND A PRIMARY  
CC DEFECT IN INSULIN SECRETION. THE CLINICAL PHENOTYPE OF MODY1 IS  
CC CHARACTERIZED BY SEVERE INSULIN SECRETORY DEFECTS, AND BY MAJOR  
CC HYPERGLYCEMIA ASSOCIATED WITH MICROVASCULAR COMPLICATIONS.  
CC -1- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.  
CC NR2 SUBFAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: X76930; CRA5248.1; -  
CC EMBL: X87870; CRA61133.1; -  
CC EMBL: X87871; CRA61134.1; -  
CC EMBL: X87872; CRA61135.1; -  
CC EMBL: Z49825; CRA89989.1; ALT-INIT.  
CC EMBL: U72369; AAB48082.1; ALT-SEQ.  
CC EMBL: U72959; AAB48082.1; JOINED.  
CC EMBL: U72961; AAB48082.1; JOINED.  
CC EMBL: U72962; AAB48082.1; JOINED.  
CC EMBL: U72963; AAB48082.1; JOINED.  
CC EMBL: U72964; AAB48082.1; JOINED.  
CC EMBL: U72965; AAB48082.1; JOINED.  
CC EMBL: U72966; AAB48082.1; JOINED.  
CC EMBL: U72967; AAB48082.1; JOINED.  
CC EMBL: U72968; AAB48082.1; JOINED.  
CC EMBL: U72969; AAB48083.1; -  
CC EMBL: U72959; AAB48083.1; JOINED.  
CC EMBL: U72961; AAB48083.1; JOINED.  
CC EMBL: U72962; AAB48083.1; JOINED.  
CC EMBL: U72963; AAB48083.1; JOINED.  
CC EMBL: U72964; AAB48083.1; JOINED.  
CC EMBL: U72965; AAB48083.1; JOINED.  
CC EMBL: U72966; AAB48083.1; JOINED.  
CC HSP; P19793.2.NLL.  
CC TRANSFAC; T00373; -  
CC TRANSFAC; T02421; -  
CC TRANSFAC; T02425; -  
CC TRANSFAC; T02428; -  
CC MIM; 600281; -  
CC MIM; 125850; -  
CC PROSITE; PS00031; NUCLEAR\_RECEPTOR; 1.  
CC PFAM; PF0104; hormone\_rec; 1.  
CC PFAM; PF0105; zf-C4; 1.  
KW Receptor; Transcription regulation; DNA-binding; Nuclear protein;  
KW Zinc-finger; Liver; Alternative splicing; Diabetes; Disease mutation;  
KW Polymorphism. 51 116 C4-TYPE ZINC FINGERS (TWO).  
FT FT C4-TYPE. 51 71 C4-TYPE.  
FT FT ZN\_FING 87 111 C4-TYPE.  
FT FT VARSPLIC 29 29 ISOFORM HNF4-ALPHA-4).  
FT FT VARSPLIC 409 419 ISOFORM HNF4-ALPHA-  
FT FT VARSPLIC 369 465 ISOFORM HNF4-ALPHA-4).  
FT FT VARSPLIC 127 127 R -> W (IN MODY1).  
FT FT VARSPLIC 130 130 /FTID=VAR\_004668.  
FT FT VARSPLIC 393 393 T -> I.  
FT FT VARSPLIC 393 393 V -> I (IN MODY1; REDUCED TRANSCRIPTION  
FT FT ACTIVITY).  
FT FT /FTID=VAR\_004670.

ID HN4A\_RAT STANDARD; PRT; 465 AA.  
AC P22449;  
DT 01-AUG-1991 (Rel. 19, Created)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE HEPATOCYTE NUCLEAR FACTOR 4-ALPHA (HNF-4-ALPHA) (TRANSCRIPTION FACTOR  
DE HNF-4) (TRANSCRIPTION FACTOR 14).  
GN HNF4A OR NR2A1 OR TCF14 OR HNF4 OR HNF-4.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
RN [1]  
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
RC TISSUE=LIVER;  
RX MEDLINE; 91122637.  
RA SLADEK F.M., ZHONG W., LAI E., DARNELL J.E. JR.;  
RT "Liver-enriched transcription factor HNF-4 is a novel member of the  
RT steroid hormone receptor superfamily.";  
RL Genes Dev. 4:2353-2365(1990).  
RN [2]  
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.  
RC STRAIN=Wistar; TISSUE=LIVER;  
RX MEDLINE; 92305063.  
RA HATA S., TSUKAMOTO T., OSUMI T.;  
RT "A novel isoform of rat hepatocyte nuclear factor 4 (HNF-4).";  
RL Biochim. Biophys. Acta 1131:211-213(1992).  
CC -1- FUNCTION: TRANSCRIPTIONALLY CONTROLLED TRANSCRIPTION FACTOR. BINDS  
CC TO DNA SITES REQUIRED FOR THE TRANSCRIPTION OF ALPHA 1-  
CC ANTIHYPSIN, APOLOPROTEIN CIII, TRANSTHYRETIN GENES AND HNF1-  
CC ALPHA. MAY BE ESSENTIAL FOR DEVELOPMENT OF THE LIVER, KIDNEY AND  
CC INTESTINE.  
CC -1- SUBUNIT: HOMODIMERIZATION IS REQUIRED FOR HNF4-ALPHA TO BIND TO  
CC ALPHA.  
CC ITS RECOGNITION SITE.  
CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
CC -1- ALTERNATIVE PRODUCTS: TWO FORMS ARE PRODUCED BY ALTERNATIVE  
CC SPLICING.  
CC -1- TISSUE SPECIFICITY: LIVER, KIDNEY AND INTESTINE.  
CC -1- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.  
CC NR2 SUBFAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; D10554; BAA01411.1; -;  
CC EMBL; X57133; CAA04012.1; -;  
CC PIR; A36471; A36471.  
CC HSP; P19793; 2NLL.  
CC TRANSFAC; T02422; -;  
CC PROSITE; PS00031; NUCLEAR\_RECEPTOR; 1.  
CC PFAM; PF00104; hormone\_rec; 1.  
CC PFAM; PF00105; zf-C4; 1.  
CC Receptor; Transcription regulation; DNA-binding; Nuclear protein;  
CC Zinc-finger; Liver; Alternative splicing.  
CC MOD\_RES 71 71 BLOCKED.  
CC C4-TYPE ZINC FINGERS (TWO).  
CC ZN\_FING 51 116 C4-TYPE.  
CC ZN\_FING 51 71 C4-TYPE.  
CC ZN\_FING 87 111 C4-TYPE.  
CC VARSPLIC 409 419 CEMPRRGQAA -> S (IN SHORT ISOFORM).  
CC CONFLICT 171 171 K -> R (IN REF. 1).  
CC CONFLICT 174 174 N -> S (IN REF. 1).  
CC SEQUENCE 465 AA; 51695 MW; A15DDFFF CRC32;  
Query Match 59.6%; Score 53; DB 1; Length 465;  
Best Local Similarity 66.7%; Pred. No. 3.76e+00;  
Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
Db 106 CRLKCCFRA 114  
| | | | |

QY 1 CILESCFRA 9  
RESULT 9  
ID HN4A\_MOUSE STANDARD; PRT; 465 AA.  
AC P49698;  
DT 01-FEB-1996 (Rel. 33, Created)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE HEPATOCYTE NUCLEAR FACTOR 4-ALPHA (HNF-4-ALPHA) (TRANSCRIPTION FACTOR  
DE HNF-4) (TRANSCRIPTION FACTOR 14).  
GN HNF4A OR NR2A1 OR TCF14 OR HNF4 OR HNF-4.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=B6/CBA; TISSUE=LIVER;  
RX MEDLINE; 95092794.  
RA HATA S., INOUE T., KOSUGA K., NAKASHIMA T., TSUKAMOTO T., OSUMI T.;  
RT "Identification of two splice isoforms of mRNA for mouse hepatocyte  
RT nuclear factor 4 (HNF-4).";  
RL Biochim. Biophys. Acta 1260:55-61(1995).  
CC -1- FUNCTION: TRANSCRIPTIONALLY CONTROLLED TRANSCRIPTION FACTOR. BINDS  
CC TO DNA SITES REQUIRED FOR THE TRANSCRIPTION OF ALPHA 1-  
CC ANTIHYPSIN, APOLOPROTEIN CIII, TRANSTHYRETIN GENES AND HNF1-  
CC ALPHA. MAY BE ESSENTIAL FOR DEVELOPMENT OF THE LIVER, KIDNEY AND  
CC INTESTINE.  
CC -1- SUBUNIT: HOMODIMERIZATION IS REQUIRED FOR HNF4-ALPHA TO BIND TO  
CC ITS RECOGNITION SITE (BY SIMILARITY).  
CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
CC -1- ALTERNATIVE PRODUCTS: TWO FORMS ARE PRODUCED BY ALTERNATIVE  
CC SPLICING.  
CC -1- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.  
CC NR2 SUBFAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; D29015; BAA06101.1; -;  
CC HSP; P19793; 2NLL.  
CC TRANSFAC; T02423; -;  
CC MGD; MGI:109128; HNF4.  
CC PROSITE; PS00031; NUCLEAR\_RECEPTOR; 1.  
CC PFAM; PF00104; hormone\_rec; 1.  
CC PFAM; PF00105; zf-C4; 1.  
CC Receptor; Transcription regulation; DNA-binding; Nuclear protein;  
CC Zinc-finger; Liver; Alternative splicing.  
CC DNA\_BIND 51 116 C4-TYPE ZINC FINGERS (TWO).  
CC ZN\_FING 51 71 C4-TYPE.  
CC ZN\_FING 51 111 C4-TYPE.  
CC VARSPLIC 409 419 CEMPRRGQAA -> S (IN SHORT ISOFORM).  
CC SEQUENCE 465 AA; 51755 MW; 097865A9 CRC32;  
Query Match 59.6%; Score 53; DB 1; Length 465;  
Best Local Similarity 66.7%; Pred. No. 3.76e+00;  
Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
Db 106 CRLKCCFRA 114  
| | | | |  
QY 1 CILESCFRA 9  
RESULT 10  
ID HN4A\_HUMAN STANDARD; PRT; 465 AA.  
AC P41235; Q92653; Q92654; Q92655; Q14540; Q99864; Q00723; Q00659;  
DT 01-FEB-1995 (Rel. 31, Created)  
DT 15-JUL-1999 (Rel. 38, Last sequence update)



```

RESULT 6
ID FSHB_RAT STANDARD; PRT; 130 AA.
AC F18427;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE FOLLITROPIN BETA CHAIN PRECURSOR (FOLLICLE-STIMULATING HORMONE)
DE (FSH-B).
GN FSHB.

OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-HOLTZMAN;
RX MEDLINE; 91042355.
RA MAURER R.A.;
RT "Molecular cloning and nucleotide sequence analysis of complementary
RT deoxyribonucleic acid for the beta-subunit of rat follicle
RT stimulating hormone.";
RL Mol. Endocrinol. 1:717-723(1987).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 89356263.
RA GHARIB S.D., ROY A., WIEMAN M.E., CHIN W.W.;
RT "Isolation and characterization of the gene encoding the beta-subunit
RT of rat follicle-stimulating hormone.";
RL DNA 8:339-349(1989).
RN [3]
RP SEQUENCE OF 55-130 FROM N.A.
RC STRAIN-SPRAGUE-DAWLEY;
RA KATO Y., EZASHI T., HIRAI T., KATO T.;
RT subunit cDNAs and gene fragment.";
RL Zool. Sci. 7:877-885(1990).
CC -1- FUNCTION: STIMULATES DEVELOPMENT OF FOLLICLE AND SPERMATOGENESIS
CC -1- IN THE REPRODUCTIVE ORGANS.
CC -1- SUBUNIT: HETERODIMER OF A COMMON ALPHA CHAIN AND A UNIQUE BETA
CC CHAIN WHICH CONFERS BIOLOGICAL SPECIFICITY TO THYROTROPIN,
CC LUTROPIN, FOLLITROPIN AND GONADOTROPIN.
CC -1- SIMILARITY: BELONGS TO THE GLYCOPROTEIN HORMONES BETA CHAIN
CC FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M35804; ; NOT_ANNOTATED_CDS.
CC EMBL; M27048; AAB60705.1; -.
CC EMBL; M27044; AAB60705.1; JOINED.
CC EMBL; D00577; BAA00455.1; -.
CC PIR; A32893; A32893.
CC PIR; A40060; A40060.
CC HSP; P01233; IHRP.
CC PROSITE; PS00261; GLYCO_HORMONE_BETA_1; FALSE_NEG.
CC DR PFAM; PS00689; GLYCO_HORMONE_BETA_2; 1.
CC DR PFAM; PF00007; Cys_knot; 1.
CC Hormone; Glycoprotein; Signal.
FT SIGNAL 1 20
FT CHAIN 21 130 FOLLITROPIN BETA CHAIN.
FT DISULFID 22 70 BY SIMILARITY.
FT DISULFID 36 85 BY SIMILARITY.
FT DISULFID 39 123 BY SIMILARITY.
FT DISULFID 47 101 BY SIMILARITY.
FT DISULFID 51 103 BY SIMILARITY.
FT DISULFID 106 113 BY SIMILARITY.
FT CARBOHYD 26 26 PROBABLE.

FT CARBOHYD 43 43 PROBABLE.
FT VARIANT 73 73 K -> R (IN STRAIN SPRAGUE-DAWLEY).
SQ SEQUENCE 130 AA; 14814 MW; 941FA1D6 CRC32;

Query Match 50.7%; Score 54; DB 1; Length 130;
Best Local Similarity 70.0%; Pred. No. 2.37e+00;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 8 CILLWCLRAV 17
||| ||||
QY 1 CILESCFRAV 10

RESULT 7
ID HN4A_XENLA STANDARD; PRT; 455 AA.
AC Q91766;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE HEPATOCYTE NUCLEAR FACTOR 4-ALPHA (HNF4-ALPHA).
DE HNF4A OR NR2A1 OR HNF4.
GN Xenopus laevis (African clawed frog).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Amphibia;
OC Batrachia; Anura; Mesobatrachia; Pipidae; Xenopodinae;
OC Xenopus.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RX MEDLINE; 96404127.
RA HOLEWA B., POGGE V., STRANDMANN E., ZAPP D., LORENZ P., RYFFEL G.U.;
RT "Transcriptional hierarchy in Xenopus embryogenesis: HNF4 a maternal
RT factor involved in the developmental activation of the gene encoding
RT the tissue specific transcription factor HNF1 alpha (LFB1).";
RL Mech. Dev. 54:45-57(1996).
CC -1- FUNCTION: TRANSCRIPTION FACTOR; BINDS AND ACTIVATES THE PROMOTER
CC FOR THE HNF1-ALPHA GENE.
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- TISSUE SPECIFICITY: EXPRESSED IN LIVER AND KIDNEY.
CC -1- DEVELOPMENTAL STAGE: EXPRESSED EARLY DURING OOGENESIS AND IS
CC ABSENT IN THE EGG.
CC -1- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
CC NR2 SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z37526; CAA85763.1; -.
CC TRANSFAC; T02429; -.
CC PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
CC PFAM; PF00104; hormone_rec; 1.
CC PFAM; PF00105; zf-C4; 1.
CC KW Receptor; Transcription regulation; DNA-binding; Nuclear protein;
CC Zinc-finger; Activator.
CC DNA_BIND 51 116 C4-TYPE ZINC FINGERS (TWO).
CC ZN_FING 51 71 C4-TYPE.
CC ZN_FING 87 111 C4-TYPE.
CC SEQUENCE 455 AA; 50938 MW; BFDC5C1D CRC32;

Query Match 59.6%; Score 53; DB 1; Length 455;
Best Local Similarity 66.7%; Pred. No. 3.76e+00;
Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Db 106 CRLKCFRA 114
||| ||||
QY 1 CILESCFRA 9

RESULT 8
```



CC -!- FUNCTION: NOT KNOWN, THOUGH MAY PLAY A ROLE IN EMBRYONAL  
CC DEVELOPMENT AND TUMOR TRANSFORMATION OR ASPECTS OF TUMOR  
CC PROGRESSION. ANTIGEN RECOGNIZED ON A MELANOMA BY AUTOLOGOUS  
CC CYTOLYTIC T LYMPHOCYTES.  
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.  
CC -!- TISSUE SPECIFICITY: EXPRESSED IN MANY TUMORS OF SEVERAL TYPES,  
CC SUCH AS MELANOMA, HEAD AND NECK SQUAMOUS CELL CARCINOMA, LUNG  
CC CARCINOMA AND BREAST CARCINOMA, BUT NOT IN NORMAL TISSUES EXCEPT  
CC FOR TESTES. NEVER EXPRESSED IN KIDNEY TUMORS, LEUKEMIAS AND  
CC LYMPHOMAS.  
CC -!- POLYMORPHISM: THE VARIANT AT POSITION 32 LIKELY REPRESENTS A  
CC POLYMORPHISM OF THE MAG-1 GENE.  
CC -!- SIMILARITY: BELONGS TO THE MAG-1 FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: M77481; AAA03229.1; -  
CC EMBL: U83672; AAB54061.1; -  
CC MIM: 300016; -  
CC PFM: PF01454; MAG-1.  
CC DR PFAM; PF01454; MAG-1.  
CC DR Antigen; Multigene family; Polymorphism; Tumor antigen.  
CC KW VARIANT 32 T -> A.  
CC FT 32 /FTID=VAR\_004283.  
CC FT DOMAIN 33 36 POLY-SER.  
CC FT MUTAGEN 163 163 D->A: ABOLISHES HLA-A1 BINDING.  
CC FT MUTAGEN 169 169 Y->A: ABOLISHES HLA-A1 BINDING.  
CC FT CONFLICT 72 72 R -> Q (IN REF. 3).  
CC SQ SEQUENCE 309 AA; 34342 MW; E5CB1300 CRC32;  
Query Match 75.3%; Score 67; DB 1; Length 309;  
Best Local Similarity 90.9%; Pred. No. 3.69e-03;  
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Db 92 CILESCFRAVI 102  
|||||  
QY 1 CILESCFRAVI 11  
-----  
RESULT 2  
ID Y045 SCHPO STANDARD; PRT; 1583 AA.  
AC Q09725;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 01-NOV-1995 (Rel. 32, Last annotation update)  
DE HYPOTHETICAL 180.2 KD PROTEIN C31A2.05C IN CHROMOSOME 1.  
GN SPAC31A2.05C.  
OS Schizosaccharomyces pombe (Fission yeast).  
OC Eukaryota; Fungi; Ascomycota; Archiascomycetes;  
OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
OC Schizosaccharomycetes.  
[1]  
RN STRAIN=972;  
RC SEQUENCE FROM N.A.  
RA DEVLIN K., CHURCHER C.M., BARRELL B.G., RAJANDREAM M.A., WALSH S.V.;  
RL Submitted (JUL-1995) to the EMBL/Genbank/DBJ databases.  
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: Z50113; CAA90463.1; -  
CC Hypothetical protein; Transmembrane.  
KW

FT TRANSMEM 319 339 POTENTIAL.  
FT TRANSMEM 633 653 POTENTIAL.  
FT TRANSMEM 764 784 POTENTIAL.  
FT TRANSMEM 1014 1034 POTENTIAL.  
FT TRANSMEM 1216 1236 POTENTIAL.  
FT TRANSMEM 1452 1472 POTENTIAL.  
SQ SEQUENCE 1583 AA; 180203 MW; 9617C75D CRC32;  
Query Match 68.5%; Score 61; DB 1; Length 1583;  
Best Local Similarity 85.7%; Pred. No. 8.08e-02;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
Db 1306 CILDSCF 1312  
|||||  
QY 1 CILESCF 7  
-----  
RESULT 3  
ID ENV\_JSRV STANDARD; PRT; 615 AA.  
AC P31621;  
DT 01-JUL-1993 (Rel. 26, Created)  
DT 01-JUL-1993 (Rel. 26, Last sequence update)  
DT 15-JUL-1998 (Rel. 36, Last annotation update)  
DE ENV POLYPROTEIN PRECURSOR (COAT POLYPROTEIN) [CONTAINS: COAT PROTEIN  
DE GP52; COAT PROTEIN GP36].  
GN ENV.  
OS Sheep pulmonary adenomatosis virus (Jaagsiekte sheep retrovirus)  
OS (JSRV).  
CC Viruses; Retroviridae; Retroviruses.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA MEDLINE; 92333675.  
RX YORK D.F., VIGNE R., VERWOERD D.W., QUERAT G.;  
RT "Nucleotide sequence of the jaagsiekte retrovirus, an exogenous and  
RT endogenous type D and B retrovirus of sheep and goats.";  
RL J. Virol. 66:4930-4939(1992).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: M80216; AAA89184.1; -  
CC PIR: E42740; VCMVJA.  
KW Coat protein; Glycoprotein; Polyprotein; Transmembrane.  
FT PROPEP 1 79 POTENTIAL.  
FT CHAIN 80 378 COAT PROTEIN GP52 (POTENTIAL).  
FT CHAIN 379 615 COAT PROTEIN GP36 (POTENTIAL).  
FT DOMAIN 80 378 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 379 402 POTENTIAL.  
FT DOMAIN 403 615 CYTOPLASMIC (POTENTIAL).  
FT CARBOHYD 108 108 POTENTIAL.  
FT CARBOHYD 127 127 POTENTIAL.  
FT CARBOHYD 178 178 POTENTIAL.  
FT CARBOHYD 219 219 POTENTIAL.  
FT CARBOHYD 275 275 POTENTIAL.  
FT CARBOHYD 319 319 POTENTIAL.  
SQ SEQUENCE 615 AA; 69343 MW; 78B74F63 CRC32;  
Query Match 65.2%; Score 58; DB 1; Length 615;  
Best Local Similarity 60.0%; Pred. No. 3.55e-01;  
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;  
Db 323 CILTNCRGV 332  
|||||  
QY 1 CILESCFRAV 10  
-----  
RESULT 4  
ID RPOA\_EAV STANDARD; PRT; 3175 AA.  
RPOA\_EAV

\*\*\*\*\*

MAGE1H (TM)

\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Fri Apr 14 23:13:22 2000; MasPar time 5.14 Seconds  
Tabular output not generated. 63.914 Million cell updates/sec

Title: >US-08-452-843-3  
Description: (1-11) from US08452843.pep  
Perfect Score: 89  
Sequence: 1 CILESCRAVI 11

Scoring table: PAM 150  
Gap 15

Searched: 82229 seqs, 29864866 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: swiss-prot38  
1:swissprot

Statistics: Mean 25.079; Variance 31.090; scale 0.807

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	67	75.3	309	1	MAGE1_HUMAN	
2	61	68.5	1593	1	YAM5_SCHPO	3.69e-03
3	58	65.2	615	1	ENV_JSVR	8.08e-02
4	57	64.0	3175	1	POL_POLYPROTEIN PRECUR	3.52e-01
5	56	62.9	427	1	YIEM_ECOLI	5.76e-01
6	54	60.7	130	1	FOLLITROPIN BETA CHAIN	9.28e-01
7	53	59.6	455	1	HNA4_XENLA	2.37e+00
8	53	59.6	465	1	HNA4_RAT	3.76e+00
9	53	59.6	465	1	HNA4_MOUSE	3.76e+00
10	53	59.6	465	1	HNA4_HUMAN	3.76e+00
11	53	59.6	665	1	NUC2_SCHPO	3.76e+00
12	53	59.6	1361	1	GLI4_XENLA	3.76e+00
13	53	59.6	137	1	CID_DROME	3.76e+00
14	52	58.4	117	1	VENOM ALLERGEN IV (ALL	3.76e+00
15	52	58.4	129	1	YIOL_YEAST	5.92e+00
16	52	58.4	489	1	CBP1_CANAL	5.92e+00
17	52	58.4	5179	1	MUC2_HUMAN	5.92e+00
18	51	57.3	446	1	HNA4_XENLA	9.27e+00
19	51	57.3	545	1	YR49_CAEEL	9.27e+00
20	51	57.3	774	1	HNA4_HUMAN	9.27e+00
21	50	56.2	219	1	TCC3_MOUSE	1.44e+01
22	50	56.2	165	1	CALL_ARATH	1.44e+01
23	50	56.2	308	1	PLSC_COCON	1.44e+01

24	50	56.2	310	1	YH29_YEAST	HYPOTHETICAL 34.1 KD P	1.44e+01
25	50	56.2	341	1	TA2R_RAT	THROMBOXANE A2 RECEPTOR	1.44e+01
26	50	56.2	341	1	TA2R_MOUSE	THROMBOXANE A2 RECEPTOR	1.44e+01
27	50	56.2	386	1	NUC2_MOUSE	NADH-UBIQUINONE OXIDOR	1.44e+01
28	50	56.2	393	1	VNS3_ROTTC	NONSTRUCTURAL RNA-BIND	1.44e+01
29	50	56.2	491	1	CGEL_MOUSE	GLI/S-SPECIFIC CYCLIN E	1.44e+01
30	50	56.2	600	1	YG48_YEAST	HYPOTHETICAL 68.3 KD P	1.44e+01
31	50	56.2	818	1	PPSA_SYNY3	PHOSPHOENOLPYRUVATE SY	1.44e+01
32	50	56.2	885	1	APCE_AGLNE	PHYCOBILISOME LINKER P	1.44e+01
33	49	55.1	58	1	TX50_DENJA	TOXIN SSC10.	2.22e+01
34	49	55.1	149	1	YANL_SCHPO	HYPOTHETICAL PROTEIN C	2.22e+01
35	49	55.1	167	1	THIM_MAIZE	THIOREDOXIN M-TYPE, CH	2.22e+01
36	49	55.1	251	1	GRC1_BACSU	PROBABLE HEPTAPENYL D	2.22e+01
37	49	55.1	260	1	PPH_MYCGE	PUTATIVE PROTEIN PHOSP	2.22e+01
38	49	55.1	279	1	Y4AC_RHISN	PUTATIVE PHYTOENE SYNT	2.22e+01
39	49	55.1	390	1	EAR2_MOUSE	ORPHAN NUCLEAR RECEPT	2.22e+01
40	49	55.1	462	1	UN47_CAEEL	UNC-47 PROTEIN.	2.22e+01
41	49	55.1	495	1	CD5_BOVIN	T-CELL SURFACE GLYCOPR	2.22e+01
42	49	55.1	955	1	TSP4_XENLA	THROMBOSPONDIN 4 PRECU	2.22e+01
43	49	55.1	956	1	TSP3_HUMAN	THROMBOSPONDIN 3 PRECU	2.22e+01
44	49	55.1	1420	1	YM8A_YEAST	HYPOTHETICAL 163.6 KD	2.22e+01
45	49	55.1	1451	1	EM30_ARATH	PATTERN FORMATION PROT	2.22e+01

ALIGNMENTS

RESULT	1	STANDARD;	PRT;	309 AA.
ID	MAGE1_HUMAN			
AC	P43355; O00346;			
DT	01-NOV-1995 (Rel. 32, Created)			
DT	01-NOV-1995 (Rel. 32, Last sequence update)			
DT	15-DEC-1999 (Rel. 39, Last annotation update)			
DE	MELANOMA-ASSOCIATED ANTIGEN 1 (MAGE-1 ANTIGEN) (ANTIGEN MZ2-E).			
GN	MAGEAL OR MAGE1 OR MAGE1A.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;			
OC	Eutheria; Primates; Catarrhini; Hominidae; Homo.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE; 92086861.			
RA	VAN DER BRUGGEN P., TRAVERSARI C., CHOMEZ P., LURQUIN C., DE PLAEN E.,			
RA	VAN DEN EYNDE B., KNUTH A., BOON T.;			
RT	"A gene encoding an antigen recognized by cytolytic T lymphocytes on			
RT	a human melanoma.";			
RL	Science 254:1643-1647(1991).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE-SKIN;			
RX	MEDLINE; 94311935.			
RA	DING M., BECK R.J., KELLER C.J., FENTON R.G.;			
RT	"Cloning and analysis of MAGE-1-related genes";			
RL	Biochem. Biophys. Res. Commun. 202:549-555(1994).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RA	GLOCKNER G., RUMP A., NORDSIEK G., HINZMANN B., KIOSCHIS P.,			
RA	HEISS N., POUSTKA A., BAUER D., DRESCHER B., KNOB A., ROSENTHAL A.;			
RL	Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.			
RN	[4]			
RP	MUTAGENESIS.			
RC	TISSUE-BLOOD;			
RX	MEDLINE; 94157413.			
RA	GAUGLER B., VAN DEN EYNDE B., VAN DER BRUGGEN P., ROMERO P.,			
RA	GAROFIO J.J., DE PLAEN E., LETHÉ B., BRASSEUR F., BOON T.;			
RT	"Human gene MAGE-3 codes for an antigen recognized on a melanoma by			
RT	autologous cytolytic T lymphocytes.";			
RL	J. Exp. Med. 179:921-930(1994).			
RN	[5]			
RP	SUBCELLULAR LOCATION.			
RX	MEDLINE; 95012905.			
RA	SCHULTZ-THATER E., JURTEIC A., DELLABONA P., LUSCHER U., SIEGRIST W.,			
RA	HARDER F., HEBERER M., ZUBER M., SPAGNOLO G.C.;			
RT	"MAGE-1 gene product is a cytoplasmic protein.";			
RL	Int. J. Cancer 59:435-439(1994).			

THIS PAGE BLANK (USPTO)

```

QY      1 CILESCFRA 9

RESULT  13
ENTRY   A30550      #type complete
TITLE   Complement C3b/C4b receptor precursor - mouse
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE    03-Mar-1989 #sequence_revision 03-Mar-1989 #text_change
12-May-1995

ACCESSIONS A30550
REFERENCE   A30550
#authors   Paul, M.S.; Aegerter, M.; O'Brien, S.E.; Kurtz, C.B.; Weis, J.H.
#journal   J. Immunol. (1989) 142:582-589
#title     The murine complement receptor gene family. Analysis of mCRY
           gene products and their homology to human CR1.
#cross-references MUID:89093944
#accession  A30550
#status    preliminary
#molecule_type mRNA
#residues  1-433 #label PAU
CLASSIFICATION #superfamily complement factor H repeat homology
FEATURE
42-98      #domain complement factor H repeat homology #label FH1\
103-160    #domain complement factor H repeat homology #label FH02\
165-231    #domain complement factor H repeat homology #label FH3\
237-293    #domain complement factor H repeat homology #label FH4\
299-355    #domain complement factor H repeat homology #label FH5
SUMMARY    #length 433 #molecular-weight 48344 #checksum 2181

Query Match      59.6%; Score 53; DB 2; Length 433;
Best Local Similarity 66.7%; Pred. NO. 1.06e+01;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db      290 LPSCFKGVI 298
QY      3 LESCFAVI 11

RESULT  14
ENTRY   S55631      #type complete
TITLE   virion protein kinase 36 - equine herpesvirus 2
ORGANISM #formal_name equine herpesvirus 2
DATE    27-Oct-1995 #sequence_revision 03-Nov-1995 #text_change
09-Sep-1997

ACCESSIONS S55631
REFERENCE   S55594
#authors   Telford, E.A.R.; Watson, M.S.; Aird, H.C.; Perry, J.;
           Davison, A.J.
#journal   J. Mol. Biol. (1995) 249:520-528
#title     The DNA sequence of equine herpesvirus 2.
#accession S55631
#status    preliminary; nucleic acid sequence not shown;
           translation not shown
#molecule_type DNA
#residues  1-438 #label TEL
#cross-references GB:U20824; NID:g695172; PID:g695209
#note      The nucleotide sequence was submitted to the EMBL Data
           Library, February 1995
SUMMARY    #length 438 #molecular-weight 49385 #checksum 5511

Query Match      59.6%; Score 53; DB 2; Length 438;
Best Local Similarity 62.5%; Pred. NO. 1.06e+01;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db      265 CVLLRCFR 272
QY      1 CILESCFR 8

RESULT  15
ENTRY   A43519      #type complete
TITLE   complement receptor CR1 precursor - mouse

```

```

ORGANISM #formal_name Mus musculus #common_name house mouse
DATE    28-Oct-1992 #sequence_revision 30-Jan-1993 #text_change
12-May-1995

ACCESSIONS A43519
REFERENCE   A43519
#authors   Paul, M.S.; Aegerter, M.; Cepek, K.; Miller, M.D.; Weis, J.H.
#journal   J. Immunol. (1990) 144:1988-1996
#title     The murine complement receptor gene family. The genomic and
           transcriptional complexity of the Cr1 and Cr2-ps genes.
#cross-references MUID:90171600
#accession  A43519
#status    preliminary
#molecule_type DNA
#residues  1-440 #label PAU
#cross-references GB:M34164
#note      the authors translated the codon GGC for residue 21 as
           Ala, and CAG for residue 121 as Glu
CLASSIFICATION #superfamily complement factor H repeat homology
FEATURE
42-98      #domain complement factor H repeat homology #label FH1\
103-160    #domain complement factor H repeat homology #label FH02\
165-231    #domain complement factor H repeat homology #label FH3\
237-293    #domain complement factor H repeat homology #label FH4\
299-355    #domain complement factor H repeat homology #label FH5
SUMMARY    #length 440 #molecular-weight 49074 #checksum 3752

Query Match      59.6%; Score 53; DB 2; Length 440;
Best Local Similarity 66.7%; Pred. NO. 1.06e+01;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db      290 LPSCFKGVI 298
QY      3 LESCFAVI 11

Search completed: Fri Apr 14 23:13:03 2000
Job time : 14 secs.

```

```

#journal      Mol. Endocrinol. (1987) 1:717-723
#title        Molecular cloning and nucleotide sequence analysis of
              complementary deoxyribonucleic acid for the beta-subunit of
              rat follicle stimulating hormone.
#cross-references MUID:91042555
#accession     A40060
##status       preliminary
##molecule_type mRNA
##residues     1-130 #label MAU
##cross-references GB:M36804
CLASSIFICATION #superfamily pituitary glycoprotein hormone beta chain
FEATURE
22-47,36-70,39-101,
51-123,85-113,
103-106
SUMMARY
#length 130 #molecular-weight 14814 #checksum 2096
Query Match 60.7%; Score 54; DB 2; Length 130;
Best Local Similarity 70.0%; Pred. No. 7.06e+00;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
Db 8 CILLWCLRAV 17
   ||| |||
QY 1 CILESCFRAV 10

RESULT 10
ENTRY
TITLE      C69102 #type complete
           DNA mismatch recognition protein Muts - Methanobacterium
           thermoautotrophicum (strain Delta H)
ORGANISM   #formal_name Methanobacterium thermoautotrophicum
DATE       05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change
05-Jun-1998
ACCESSIONS C69102
REFERENCE   A69000
#authors    Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.;
           Dubois, J.; Aldridge, T.; Bashirzadeh, R.; Blakely, D.;
           Cook, R.; Gilbert, K.; Harrison, D.; Hoang, L.; Keagle, P.;
           Lumm, W.; Pothier, B.; Qiu, D.; Spadafora, R.; Vicaire, R.;
           Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiواني, N.; Caruso,
           A.; Bush, D.; Safer, H.; Patwell, D.; Prabhakar, S.;
           McDougall, S.; Shimer, G.; Goyal, A.; Pietrokowski, S.;
           Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling,
           J.; Reeve, J.N.
#journal    J. Bacteriol. (1997) 179:7135-7155
#title      Complete genome sequence of Methanobacterium
           thermoautotrophicum Delta H: functional analysis and
           comparative genomics.
#cross-references MUID:98037514
#accession  C69102
##status    preliminary; nucleic acid sequence not shown;
           translation not shown
##molecule_type DNA
##residues  1-647 #label MTH
##cross-references GB:AE000931; GB:AE000666; NID:g2622885; PID:g2622891
##experimental_source strain Delta H
GENETICS
#gene       MTH1762
#start_codon TTG
SUMMARY
#length 647 #molecular-weight 73592 #checksum 610
Query Match 60.7%; Score 54; DB 2; Length 647;
Best Local Similarity 60.0%; Pred. No. 7.06e+00;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
Db 487 CALEACVRVV 496
   ||| |||
QY 1 CILESCFRAV 10

RESULT 11
ENTRY
TITLE      I55975 #type fragment
           X/Y protein - mouse (fragment)

```

```

ORGANISM      #formal_name Mus musculus #common_name house mouse
DATE          26-Jul-1996 #sequence_revision 26-Jul-1996 #text_change
10-Oct-1997
ACCESSIONS    I55975
REFERENCE     Aegarter-Shaw, M.; Cole, J.L.; Klickstein, L.B.; Wong, W.W.;
              Fearon, D.T.; Lallely, P.A.; Weis, J.H.
              J. Immunol. (1987) 138:3488-3494
#journal      Expansion of the complement receptor gene family:
              Identification in the mouse of two new genes related to the
              CRI and CR2 gene family.
#cross-references MUID:87196375
#accession    I55975
##status      preliminary; translated from GB/EMBL/DBJ
##molecule_type mRNA
##residues    1-330 #label RES
##cross-references GB:M16179; NID:g202427; PID:g202428
CLASSIFICATION #superfamily complement factor H repeat homology
FEATURE
36-92         #domain complement factor H repeat homology #label FH4\
188-244       #domain complement factor H repeat homology #label FH1\
249-306       #domain complement factor H repeat homology #label FH02
SUMMARY       #length 330 #checksum 9931
Query Match 59.6%; Score 53; DB 2; Length 330;
Best Local Similarity 66.7%; Pred. No. 1.06e+01;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Db 89 LPSCFKGVI 97
   ||||| |||
QY 3 LESCPRAVI 11

RESULT 12
ENTRY
TITLE      JC4938 #type complete
           hepatocyte nuclear factor 4C - human
ORGANISM    #formal_name Homo sapiens #common_name man
DATE        22-Oct-1996 #sequence_revision 01-Nov-1996 #text_change
31-Oct-1997
ACCESSIONS  JC4938
REFERENCE   JC4936
#authors    Kritsis, A.A.; Argyrokastritis, A.; Moschonas, N.K.; Power,
           S.; Katrakili, N.; Zannis, V.I.; Cereghini, S.; Talianidis,
           I.
#journal     Gene (1996) 173:275-280
#title      Isolation and characterization of a third isoform of human
           hepatocyte nuclear factor 4.
#cross-references MUID:97082982
#accession  JC4938
##status    preliminary
##molecule_type mRNA
##residues  1-408 #label KRI
##cross-references EMBL:X87872; NID:g1595753; PID:e184046; PID:g1595754
##experimental_source liver
##note       This protein is one of the positive regulators of liver-specific
           genes.
GENETICS
#gene       hHNF-4C
#map_position 20
CLASSIFICATION #superfamily unassigned erba-related proteins; erba
              transforming protein homology
KEYWORDS     zinc finger
FEATURE
49-288       #domain erba transforming protein homology #label ERBA
SUMMARY      #length 408 #molecular-weight 45578 #checksum 5015
Query Match 59.6%; Score 53; DB 2; Length 408;
Best Local Similarity 66.7%; Pred. No. 1.06e+01;
Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Db 106 CRLKCFRA 114
   ||| |||

```

```
ORGANISM #formal_name equine arteritis virus
#note host Equus caballus (domestic horse)
DATE 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change
29-May-1998
ACCESSIONS A39925; S10158; B39925
REFERENCE #authors Den Boon, J.A.; Snijder, E.J.; Chirnside, E.D.; De Vries,
A.A.F.; Horzinek, M.C.; Spaan, W.J.M.
#journal J. Virol. (1991) 65:2910-2920
#title Equine arteritis virus is not a togavirus but belongs to the
coronavirusslike superfamily.
#cross-references MUID:91237805
#accession A39925
#molecule_type genomic RNA
#residues 1-3175 #label DEN
#cross-references EMBL:X53459
#note a -1 ribosomal frameshift occurs between the codons AAC
for 1727-Asn and CUG for 1728-Leu
REFERENCE S10158
#authors de Vries, A.A.F.; Chirnside, E.D.; Bredendbeek, P.J.;
Gravestien, L.A.; Horzinek, M.C.; Spaan, W.J.M.
#journal Nucleic Acids Res. (1990) 18:3241-3247
#title All subgenomic mRNAs of equine arteritis virus contain a
common leader sequence.
#cross-references MUID:90287699
#accession S10158
#status translation not shown
#molecule_type genomic RNA
#residues 1-17 #label VRI
#cross-references EMBL:X52277
CLASSIFICATION #superfamily equine arteritis virus RNA-directed RNA
polymerase
KEYWORDS nucleotidyltransferase
SUMMARY #length 3175 #molecular-weight 345277 #checksum 9571
Query Match 64.0%; Score 57; DB 1; Length 3175;
Best Local Similarity 60.0%; Pred. No. 2.01e+00;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
Db 344 CLDESCFRGI 353
1:|||||:
QY 1 CILESCFRAV 10
RESULT 7
ENTRY QBCO3 #type complete
TITLE hypothetical 49.6 kb protein in asna 3' region - Escherichia
coli (strain K-12)
ORGANISM #formal_name Escherichia coli
DATE 17-May-1985 #sequence_revision 30-Sep-1997 #text_change
14-Nov-1997
ACCESSIONS B65178; A04443
REFERENCE A64720
#authors Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.;
Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.;
Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.;
Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao,
Y.
#journal Science (1997) 277:1453-1462
#title The complete genome sequence of Escherichia coli K-12.
#cross-references MUID:97426617
#accession B65178
#status nucleic acid sequence not shown; translation not shown
#molecule_type DNA
#residues 1-427 #label BLAT
#cross-references GB:A5000451; GB:U00096; NID:g2367272; PID:g2367274;
URGP:b3745
#experimental_source strain K-12, substrain MG1655
REFERENCE A91504
#authors Buhr, H.J.; Messer, W.
#journal Gene (1983) 24:285-279
#title The replication origin region of Escherichia coli: nucleotide
sequence and functional units.
```

```
#cross-references MUID:84059088
#accession A04443
#molecule_type DNA
#residues 128-427 #label BUH
GENETICS
#map_position 84 min
SUMMARY #length 427 #molecular-weight 49625 #checksum 4786
Query Match 62.9%; Score 56; DB 1; Length 427;
Best Local Similarity 66.7%; Pred. No. 3.07e+00;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Db 339 LASCRAIM 347
1:|||||:
QY 3 LESCRAVI 11
RESULT 8
ENTRY T02437 #type complete
TITLE hypothetical protein T26J13.1 - Arabidopsis thaliana
ORGANISM #formal_name Arabidopsis thaliana #common_name mouse-ear
cress
DATE 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change
05-Mar-1999
ACCESSIONS T02437
REFERENCE Z14192
#authors Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.;
Brandon, R.C.; Sykes, S.M.; Kaul, S.; Mason, T.M.;
Kerlavage, A.R.; Adams, M.D.; Somerville, C.R.; Venter,
J.C.
#submission submitted to the EMBL Data Library, June 1998
#description Arabidopsis thaliana chromosome II BAC T26J13 genomic
sequence.
#accession T02437
#status preliminary; translated from GB/EMBL/DDBJ
#molecule_type DNA
#residues 1-98 #label ROU
#cross-references EMBL:AC004625; NID:g3241939; PID:g3241940
GENETICS
#map_position II
#introns 16/1; 84/2
#note T26J13.1
SUMMARY #length 98 #molecular-weight 10728 #checksum 4499
Query Match 60.7%; Score 54; DB 2; Length 98;
Best Local Similarity 57.1%; Pred. No. 7.06e+00;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
Db 92 CLDCAF 98
1:||||
QY 1 CILESCF 7
RESULT 9
ENTRY A32893 #type complete
TITLE follitropin beta chain precursor - rat
ORGANISM #formal_name Rattus norvegicus #common_name Norway rat
DATE 29-Jan-1990 #sequence_revision 29-Jan-1990 #text_change
08-Sep-1997
ACCESSIONS A32893; A40060
REFERENCE A32893
#authors Garib, S.D.; Roy, A.; Wierman, M.E.; Chin, W.W.
#journal DNA (1989) 8:339-349
#title Isolation and characterization of the gene encoding the
beta-subunit of rat follicle-stimulating hormone.
#cross-references MUID:89356263
#accession A32893
#status preliminary
#molecule_type DNA
#residues 1-130 #label GHA
#cross-references GB:M27044; GB:M27048; NID:g204179; PID:g204181
REFERENCE A40060
#authors Maurer, R.A.
```



##cross-references EMBL:Z50113; NID:g914878; PID:g914883  
GENETICS  
#introns 33/1: 98/2; 543/3; 699/3; 1294/2; 1339/3; 1558/3  
SUMMARY #length 1583 #molecular-weight 180202 #checksum 4709  
Query Match 68.5%; Score 61; DB 2; Length 1583;  
Best Local Similarity 85.7%; Pred. No. 3.54e-01;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
Db 1306 CILDSCF 1312  
|||:||||  
QY 1 CILESCF 7

RESULT 3 VCMVJA #type complete  
ENTRY env polyprotein precursor - sheep pulmonary adenomatosis  
TITLE virus  
ALTERNATE\_NAMES coat polyprotein  
CONTAINS coat protein gp36; coat protein gp52  
ORGANISM #formal\_name sheep pulmonary adenomatosis virus  
DATE 31-Mar-1993 #sequence\_revision 31-Mar-1993 #text\_change  
05-Sep-1997  
ACCESSIONS E42740  
REFERENCE A42740  
#authors York, D.F.; Vigne, R.; Verwoerd, D.W.; Querat, G.  
#journal J. Virol. (1992) 66:4930-4939  
#title Nucleotide sequence of the jaagsiekte retrovirus, an  
exogenous and endogenous type D and B retrovirus of sheep  
and goats.

##cross-references MUID:92333675  
#accession E42740  
#molecule\_type genomic RNA  
#residues 1-615 #label YOR  
##cross-references GB:M80216; NID:g331338; PID:g331342

GENETICS  
#gene env  
CLASSIFICATION #superfamily type A retrovirus env polyprotein  
KEYWORDS coat protein; glycoprotein; polyprotein; transmembrane  
protein  
FEATURE  
1-79 #domain signal sequence #status predicted #label SIG\  
80-378 #product coat protein gp52 #status predicted #label CP1\  
379-615 #product coat protein gp36 #status predicted #label CP2\  
379-402 #domain transmembrane #status predicted #label TM1\  
403-615 #domain intracellular #status predicted #label INT\  
555-571 #domain transmembrane #status predicted #label TM2\  
108,127,178,219, #binding\_site carbohydrate (Asn) (covalent) #status  
275 predicted  
SUMMARY #length 615 #molecular-weight 69343 #checksum 8020  
Query Match 65.2%; Score 58; DB 1; Length 615;  
Best Local Similarity 60.0%; Pred. No. 1.31e-00;  
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;  
Db 323 CILTCIRGV 332  
|||:||||  
QY 1 CILESCFV 10

RESULT 4 S72705 #type complete  
ENTRY mycroceroic acid synthase masA - Mycobacterium leprae  
TITLE Lepb1170\_C2\_209 protein  
ALTERNATE\_NAMES  
ORGANISM #formal\_name Mycobacterium leprae  
DATE 19-Mar-1997 #sequence\_revision 25-Apr-1997 #text\_change  
05-Mar-1999  
ACCESSIONS S72705  
REFERENCE S72693  
#authors Smith, D.R.; Robison, K.  
#submission submitted to the EMBL Data Library, November 1993  
#description Mycobacterium leprae cosmid B1170.

##accession S72705  
#status preliminary  
#molecule\_type DNA  
#residues 1-2118 #label SMI  
##cross-references EMBL:U00010; NID:g466780; PID:g466793  
GENETICS  
#start\_codon TTG  
CLASSIFICATION #superfamily mycroceroic acid synthase;  
3-oxoacyl-[acyl-carrier-protein] synthase I homology; acyl  
carrier protein homology; long-chain alcohol dehydrogenase  
homology; short-chain alcohol dehydrogenase homology;  
[acyl-carrier-protein] S-malonyltransferase homology  
FEATURE  
28-426 #domain 3-oxoacyl-[acyl-carrier-protein] synthase I  
homology #label OAS\  
536-816 #domain [acyl-carrier-protein] S-malonyltransferase  
homology #label AMT\  
1449-1738 #domain long-chain alcohol dehydrogenase homology #label  
LADH\  
1770-1954 #domain short-chain alcohol dehydrogenase homology  
#label SADH\  
2038-2110 #domain acyl carrier protein homology #label ACPI  
SUMMARY #length 2118 #molecular-weight 226495 #checksum 3824  
Query Match 65.2%; Score 58; DB 2; Length 2118;  
Best Local Similarity 50.0%; Pred. No. 1.31e-00;  
Matches 5; Conservative 5; Mismatches 0; Indels 0; Gaps 0;  
Db 1099 LLDACFSQVI 1108  
|||:||||  
QY 2 ILESCFRAVI 11

RESULT 5  
ENTRY G70384 #type complete  
TITLE hypothetical protein aq\_978 - Aquifex aeolicus  
ORGANISM #formal\_name Aquifex aeolicus  
DATE 08-May-1998 #sequence\_revision 08-May-1998 #text\_change  
08-May-1998  
ACCESSIONS G70384  
REFERENCE A70300  
#authors Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.;  
Lenox, A.L.; Graham, D.E.; Overbeek, R.; Snead, M.A.;  
Keller, M.; Aujay, M.; Huber, R.; Feldman, R.A.; Short,  
J.M.; Olson, G.J.; Swanson, R.V.  
#journal Nature (1998) 392:353-358  
#title The complete genome of the hyperthermophilic bacterium  
Aquifex aeolicus.  
##cross-references MUID:98196666  
#accession G70384  
#status preliminary; nucleic acid sequence not shown;  
translation not shown  
#molecule\_type DNA  
#residues 1-145 #label AOF  
##cross-references GB:AE000716; NID:g2983478; PID:g2983489; GB:AE000657  
##experimental\_source strain VF5  
GENETICS  
#gene aq\_978  
SUMMARY #length 145 #molecular-weight 16638 #checksum 9196  
Query Match 64.0%; Score 57; DB 2; Length 145;  
Best Local Similarity 62.5%; Pred. No. 2.01e-00;  
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
Db 9 CILECCYR 16  
|||:||||  
QY 1 CILESCFR 8

RESULT 6  
ENTRY RRW5V #type complete  
TITLE genome polyprotein - equine arteritis virus  
CONTAINS RNA-directed RNA polymerase (EC 2.7.48)

\*\*\*\*\*

W P S R E H (TM)

\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Fri Apr 14 23:12:49 2000; MasPar time 3.45 Seconds  
Tabular output not generated. 127.598 Million cell updates/sec

Title: >US-08-452-843-3  
Description: (1-11) from US08452843.pep  
Perfect Score: 89  
Sequence: 1 CILESCFRAVI 11

Scoring table: PAM 150  
Gap 15

Searched: 122810 seqs, 40068593 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: pir62  
1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 24.270; Variance 35.035; scale 0.693

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	67	75.3	280	JC2358	tumor-associated anti	2.33e-02
2	61	68.5	1583	JC2358	hypothetical protein	3.54e-01
3	58	65.2	615	1 VCMVJA	env polyprotein precu	1.31e-00
4	58	65.2	2118	2 S72705	mycotoxin acid synt	1.31e-00
5	57	64.0	145	2 G70384	hypothetical protein	2.01e-00
6	57	64.0	3175	1 RWVEV	genome polyprotein -	2.01e-00
7	56	62.9	427	1 QOEC03	hypothetical 49.6 kD	7.06e-00
8	54	60.7	98	2 T02437	hypothetical protein	7.06e-00
9	54	60.7	130	2 A32893	foliitropin beta chai	7.06e-00
10	54	60.7	647	2 C59102	DNA mismatch recognit	7.06e-00
11	53	59.6	330	2 I55975	X/Y protein - mouse	1.06e-01
12	53	59.6	408	2 J04938	hepatocyte nuclear fa	1.06e-01
13	53	59.6	433	2 A30550	complement C3b/C4b re	1.06e-01
14	53	59.6	438	2 S5631	virion protein kinase	1.06e-01
15	53	59.6	440	2 A35519	complement receptor C	1.06e-01
16	53	59.6	455	2 J04936	hepatocyte nuclear fa	1.06e-01
17	53	59.6	455	2 A36471	transcription factor	1.06e-01
18	53	59.6	463	2 J04009	hepatocyte nuclear fa	1.06e-01
19	53	59.6	465	2 J04937	hepatocyte nuclear fa	1.06e-01
20	53	59.6	465	2 S2074	hepatocyte nuclear fa	1.06e-01
21	53	59.6	465	2 S23502	hepatocyte nuclear fa	1.06e-01
22	53	59.6	504	2 J06096	hepatocyte nuclear fa	1.06e-01
23	53	59.6	605	2 D71318	probable DNA primase	1.06e-01

24	53	59.6	1377	2 A38926	DNA-binding protein c	1.06e-01
25	53	59.6	1582	2 E70876	probable polyketidesy	1.06e-01
26	52	58.4	117	2 C37330	venom allergen IV - r	1.59e-01
27	52	58.4	127	2 D69186	hypothetical protein	1.59e-01
28	52	58.4	129	2 S48393	probable membrane pro	1.59e-01
29	52	58.4	135	2 H69202	hypothetical protein	1.59e-01
30	52	58.4	273	2 D71436	hypothetical protein	1.59e-01
31	52	58.4	489	2 A47259	corticosteroid-bindin	1.59e-01
32	52	58.4	945	3 T00024	ent-raurene synthase	1.59e-01
33	52	58.4	2467	2 D71437	probable resistance g	1.59e-01
34	52	58.4	3020	2 A43932	mucin 2 precursor, in	1.59e-01
35	51	57.3	177	2 JC5748	coronafacic acid synt	2.37e-01
36	51	57.3	521	2 F64522	conserved hypothetica	2.37e-01
37	51	57.3	523	2 E71985	hypothetical protein	2.37e-01
38	51	57.3	774	2 J06095	hepatocyte nuclear fa	2.37e-01
39	51	57.3	1747	2 A45974	collagen alpha 1(XIV)	2.37e-01
40	51	57.3	2108	2 H70819	probable polyketide s	2.37e-01
41	51	57.3	2111	2 A70668	mycotoxin acid synt	2.37e-01
42	51	57.3	2126	2 E70522	probable polyketide s	2.37e-01
43	50	56.2	147	2 S37485	gene msgi protein - m	3.52e-01
44	50	56.2	169	1 RWMSC2	T-cell receptor gamma	3.52e-01
45	50	56.2	810	2 E71550	probable phospholipas	3.52e-01

ALIGNMENTS

RESULT 1 JC2358 #type complete  
ENTRY tumor-associated antigen MAGE-1 - human  
TITLE #formal\_name Homo sapiens #common\_name man  
ORGANISM 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change  
DATE 04-Sep-1998  
ACCESSIONS JC2358  
REFERENCE JC2358  
#authors Ding, M.; Beck, R.J.; Keller, C.J.; Fenton, R.G.  
#journal Biochem. Biophys. Res. Commun. (1994) 202:549-555  
#title Cloning and analysis of MAGE-1-related genes.  
#cross-references MUID:94311935  
#accession JC2358  
#molecule\_type mRNA  
##residues 1-280 ##label DIN  
##experimental\_source melanoma cell line DM150

GENETICS #gene MAGE

CLASSIFICATION #superfamily tumor associated protein MAGE

FEATURE 161-169 #region HLA-A1 binding #status predicted

SUMMARY #length 280 #molecular-weight 30932 #checksum 467

Query Match 75.3%; Score 67; DB 2; Length 280;  
Best Local Similarity 90.9%; Pred. No. 2.33e-02;  
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 92 CILESCFRAVI 102

QY 1 CILESCFRAVI 11

|||||

RESULT 2 S59644 #type complete

ENTRY hypothetical protein SPAC31A2.05c - fission yeast

TITLE (Schizosaccharomyces pombe)

ORGANISM #formal\_name Schizosaccharomyces pombe

DATE 14-Jan-1996 #sequence\_revision 19-Apr-1996 #text\_change

ACCESSIONS S59644

REFERENCE S58093

#authors Devlin, K.; Churcher, C.M.

#submission submitted to the EMBL Data Library, July 1995

#accession S59644

##status preliminary

##molecule\_type DNA

##residues 1-1583 ##label DEV

CC (see also V52730) and the MODY1 locus is the HNF-4 alpha gene (see  
 CC also V52687). Analysis of mutations in these HNF genes can be  
 CC diagnostic for diabetes. The invention also contemplates methods  
 CC of screening for modulators of HNF function utilizing HNF nucleic  
 CC acids or polypeptides, the modulators being useful for treating  
 CC diabetes by modulating HNF function in an animal.  
 SQ Sequence 567 AA;

Query Match 59.6%; Score 53; DB 1; Length 567;  
 Best Local Similarity 66.7%; Pred. No. 7.99e+01;  
 Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 Db 166 CRKLCFRA 174  
 | | | | |  
 QY 1 CILESCFRA 9

RESULT 15  
 ID R44295 standard; Protein; 489 AA.  
 AC R44295;  
 DT 28-JUN-1994 (first entry)  
 DE Corticosterone-binding protein.  
 KW Corticosterone-binding protein; hormone-binding protein; diagnosis;  
 KW therapy; steroid-binding protein; ss.  
 OS Candida albicans.  
 PN WO9324516-A.  
 PD 09-DEC-1993.  
 PF 27-MAY-1993; U05047.  
 PR 28-MAY-1992; US-890440.  
 PA (STRD ) UNIV LELAND STANFORD JUNIOR.  
 PI Feldman D, Malloy PJ;  
 DR WPI; 93-405722/50.  
 PT Nucleic acid encoding hormone-binding protein - used for  
 PT detecting organisms and for developing prods. for use in  
 PT diagnosis and therapy  
 PS Claim 2; Fig. 2; 55pp; English.  
 CC The CBP protein may be expressed recombinantly in a host  
 CC microorganism. CBP may prove to be a novel target for the  
 CC development of new therapeutic agents and diagnostic methods for  
 CC the treatment and management of candidiasis and other diseases.  
 SQ Sequence 489 AA;

Query Match 58.4%; Score 52; DB 1; Length 489;  
 Best Local Similarity 54.5%; Pred. No. 1.02e+02;  
 Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;  
 Db 472 CILENIFRNDV 482  
 | | | | |  
 QY 1 CILESCFRAV 11

Search completed: Fri Apr 14 23:12:31 2000  
 Job time : 40 secs.

Db 106 CRLKCCFRA 114  
| | | | |  
QY 1 CILESCFRA 9

RESULT 12  
ID W71574 standard; Protein: 465 AA.

AC W71574;  
DT 21-DEC-1998 (first entry)  
DE Human native hepatocyte nuclear factor 4 alpha.  
KW Hepatocyte nuclear factor 4 alpha; HNF-4 alpha; MODY1; human;  
transcription factor; maturity onset diabetes of the young;  
KS diabetes; NIDDM; diagnosis; therapy.  
OS Homo sapiens.  
PN WO9811254-A1.  
PD 19-MAR-1998.  
PF 10-SEP-1997; U16037.  
PR 30-OCT-1996; US-029679.  
PR 10-SEP-1996; US-025719.  
PR 02-OCT-1996; US-028056.  
PA (ARCH-) ARCH DEV CORP.

PI Bell GI, Furuta H, Horikawa Y, Kaisaki PJ, Menzel S,  
PI Oda N, Yamagata K;  
DR N-PSDB; V52731.  
DR 10-SEP-1997; U16037.  
DR 30-OCT-1996; US-029679.  
DR 10-SEP-1996; US-025719.  
DR 02-OCT-1996; US-028056.  
DR (ARCH-) ARCH DEV CORP.

PI Isolated nucleic acid encoding hepatocyte nuclear factor 1-alpha and  
PT 1-beta - useful for detecting susceptibility for non-insulin  
PT dependent diabetes, especially maturity-onset diabetes of the young  
PS Claim 53; Page 210-211; 363pp; English.

CC This is the amino acid sequence of human hepatocyte nuclear  
CC factor-4 alpha (HNF-4 alpha), a transcription factor involved in  
CC regulating gene expression in insulin-secreting beta cells. A cDNA  
CC sequence (see V52687) encoding HNF-4 alpha is provided. Mutations  
CC in HNF-4 alpha are indicative of a propensity to MODY1 (maturity  
CC onset diabetes of the young) type diabetes. The HNF-4 alpha gene  
CC is located on human chromosome 20, which is the location site of  
CC the MODY1 locus. The invention concerns the identification of  
CC genes responsible for non-insulin dependent diabetes mellitus  
CC (NIDDM) for use in diagnostics and therapeutics. It demonstrates  
CC that the MODY3 locus is the HNF-1 alpha gene. The MODY4 locus is  
CC the HNF-1 beta gene (see V52730) and the MODY1 locus is the HNF-4  
CC alpha gene (see V52687). Analysis of mutations in these HNF genes  
CC can be diagnostic for diabetes. The invention also contemplates  
CC methods of screening for modulators of HNF function utilising HNF  
CC nucleic acids or polypeptides, the modulators being useful for  
CC treating diabetes by modulating HNF function in an animal.  
SQ Sequence 465 AA;

Query Match 59.6%; Score 53; DB 1; Length 465;  
Best Local Similarity 66.7%; Pred. No. 7.99e+01;  
Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Db 106 CRLKCCFRA 114  
| | | | |  
QY 1 CILESCFRA 9

RESULT 13  
ID W71587 standard; Protein: 516 AA.

AC W71587;  
DT 21-DEC-1998 (first entry)  
DE Human hepatocyte nuclear factor 4 alpha.  
KW Hepatocyte nuclear factor 4 alpha; HNF-4 alpha; MODY1; human;  
transcription factor; maturity onset diabetes of the young;  
KS diabetes; NIDDM; diagnosis; therapy.  
OS Homo sapiens.  
PN WO9811254-A1.  
PD 19-MAR-1998.  
PF 10-SEP-1997; U16037.  
PR 30-OCT-1996; US-029679.  
PR 10-SEP-1996; US-025719.  
PR 02-OCT-1996; US-028056.  
PA (ARCH-) ARCH DEV CORP.

PI Isolated nucleic acid encoding hepatocyte nuclear factor 1-alpha and  
PT 1-beta - useful for detecting susceptibility for non-insulin  
PT dependent diabetes, especially maturity-onset diabetes of the young  
PS Disclosure; Fig 28A-V; 363pp; English.  
CC This is the amino acid sequence of human hepatocyte nuclear factor  
CC 4 alpha (HNF-4 alpha) as deduced from a partial gene sequence (see  
CC V52731). Alternative splicing results in a 516-amino acid  
CC polypeptide (see W71587). The HNF-4 alpha sequence has also  
CC been deduced from a cDNA clone (see W71574). HNF-4 alpha is a  
CC transcription factor involved in regulating gene expression in  
CC insulin-secreting beta cells. Mutations in HNF-4 alpha are  
CC indicative of a propensity to diabetes mellitus. The invention  
CC concerns the identification of genes responsible for non-insulin  
CC dependent diabetes mellitus (NIDDM) for use in diagnostics and  
CC therapeutics. It demonstrates that the MODY3 locus is the HNF-1  
CC alpha gene (see V52625), the MODY4 locus is the HNF-1 beta gene

PI Bell GI, Furuta H, Horikawa Y, Kaisaki PJ, Menzel S,  
PI Oda N, Yamagata K;  
DR WPI; 98-271667/24.  
DR N-PSDB; V52731.

PT Isolated nucleic acid encoding hepatocyte nuclear factor 1-alpha and  
PT 1-beta - useful for detecting susceptibility for non-insulin  
PT dependent diabetes, especially maturity-onset diabetes of the young  
PS Disclosure; Fig 28A-V; 363pp; English.

CC This is the amino acid sequence of human hepatocyte nuclear factor  
CC 4 alpha (HNF-4 alpha) as deduced from a partial gene sequence (see  
CC V52731). Alternative splicing results in a 567-amino acid  
CC polypeptide (see W71582). The HNF-4 alpha sequence has also  
CC been deduced from a cDNA clone (see W71574). HNF-4 alpha is a  
CC transcription factor involved in regulating gene expression in  
CC insulin-secreting beta cells. Mutations in HNF-4 alpha are  
CC indicative of a propensity to diabetes mellitus. The invention  
CC concerns the identification of genes responsible for non-insulin  
CC dependent diabetes mellitus (NIDDM) for use in diagnostics and  
CC therapeutics. It demonstrates that the MODY3 locus is the HNF-1  
CC alpha gene (see V52625), the MODY4 locus is the HNF-1 beta gene  
CC (see also V52730) and the MODY1 locus is the HNF-4 alpha gene (see  
CC also V52687). Analysis of mutations in these HNF genes can be  
CC diagnostic for diabetes. The invention also contemplates methods  
CC of screening for modulators of HNF function utilising HNF nucleic  
CC acids or polypeptides, the modulators being useful for treating  
CC diabetes by modulating HNF function in an animal.  
SQ Sequence 516 AA;

Query Match 59.6%; Score 53; DB 1; Length 516;  
Best Local Similarity 66.7%; Pred. No. 7.99e+01;  
Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Db 157 CRLKCCFRA 165  
| | | | |  
QY 1 CILESCFRA 9

RESULT 14  
ID W71582 standard; Protein: 567 AA.

AC W71582;  
DT 21-DEC-1998 (first entry)  
DE Human hepatocyte nuclear factor 4 alpha.  
KW Hepatocyte nuclear factor 4 alpha; HNF-4 alpha; MODY1; human;  
transcription factor; maturity onset diabetes of the young;  
KS diabetes; NIDDM; diagnosis; therapy.  
OS Homo sapiens.  
PN WO9811254-A1.  
PD 19-MAR-1998.  
PF 10-SEP-1997; U16037.  
PR 30-OCT-1996; US-029679.  
PR 10-SEP-1996; US-025719.  
PR 02-OCT-1996; US-028056.  
PA (ARCH-) ARCH DEV CORP.

PI Bell GI, Furuta H, Horikawa Y, Kaisaki PJ, Menzel S,  
PI Oda N, Yamagata K;  
DR WPI; 98-271667/24.  
DR N-PSDB; V52731.

PT Isolated nucleic acid encoding hepatocyte nuclear factor 1-alpha and  
PT 1-beta - useful for detecting susceptibility for non-insulin  
PT dependent diabetes, especially maturity-onset diabetes of the young  
PS Disclosure; Fig 28A-V; 363pp; English.

CC This is the amino acid sequence of human hepatocyte nuclear factor  
CC 4 alpha (HNF-4 alpha) as deduced from a partial gene sequence (see  
CC V52731). Alternative splicing results in a 516-amino acid  
CC polypeptide (see W71587). The HNF-4 alpha sequence has also  
CC been deduced from a cDNA clone (see W71574). HNF-4 alpha is a  
CC transcription factor involved in regulating gene expression in  
CC insulin-secreting beta cells. Mutations in HNF-4 alpha are  
CC indicative of a propensity to diabetes mellitus. The invention  
CC concerns the identification of genes responsible for non-insulin  
CC dependent diabetes mellitus (NIDDM) for use in diagnostics and  
CC therapeutics. It demonstrates that the MODY3 locus is the HNF-1  
CC alpha gene (see V52625), the MODY4 locus is the HNF-1 beta gene

AC R94701;  
 DT 04-AUG-1996 (first entry)  
 DE PRRSV VR 2385 ORF-5 product.  
 KW PRRSV; vaccine; antigen.  
 OS Pig reproductive and respiratory syndrome virus Iowa strain ISU-12.  
 PN W09606619-A1.  
 PD 07-MAR-1996.  
 PR 01-SEP-1995; U10904.  
 PR 01-SEP-1994; US-301435.  
 PA (HALB/) HALBUR P.  
 PA (LUMB/) LUM M A.  
 PA (MENG/) MENG X.  
 PA (MORO/) MOROZOV I.  
 PA (PAUL/) PAUL P S.  
 PI Halbur P, Lum MA, Meng X, Morozov I, Paul PS;  
 DR WPI: 96-160132/15.  
 DR N-PSDB; T14390.  
 PT New porcine reproductive and respiratory syndrome virus DNA - and  
 PT proteins encoded by open reading frames of an Iowa strain of the  
 PT virus; are used in vaccines against PRRSV in pigs  
 PS Disclosure: Page 140-141; 228pp; English.  
 CC The protein (R94701) encoded by open reading frame 5 (ORF-5 -  
 CC T14390) of porcine reproductive and respiratory syndrome virus  
 CC (PRRSV) Iowa strain isolate ISU-12 (VR 2385) can be used with  
 CC R94719-21) in the development of vaccines against PRRSV in pigs  
 CC and in serological tests for screening pigs for exposure to, or  
 CC infection by, PRRSV (partic. strain Iowa).  
 SQ Sequence 200 AA;  
 Query Match 59.6%; Score 53; DB 1; Length 200;  
 Best Local Similarity 54.5%; Pred. No. 7.99e+01;  
 Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;  
 Db 19 CIVPSCFVALV 29  
 QY 1 CILESCFRAVI 11  
 ||: ||| |::  
 RESULT 10  
 ID W25953 standard; Protein; 200 AA.  
 AC W25953;  
 DT 10-NOV-1997 (first entry)  
 DE ORF 5 protein of PRRSV isolate VR2385.  
 KW Porcine reproductive and respiratory syndrome virus; PRRSV; coronavirus;  
 KW reproductive failure; pneumonia; pig; preweaning mortality; torovirus;  
 KW subgenomic mRNA; glycosylated membrane protein; nucleocapsid protein;  
 KW membrane associated protein; vaccine; antibody; therapy.  
 OS Porcine reproductive and respiratory syndrome virus.  
 PN W09640932-A1.  
 PD 19-DEC-1996.  
 PR 07-JUN-1995; U08962.  
 PR 07-JUN-1995; US-478316.  
 PA (HALB/) HALBUR P.  
 PA (MENG/) MENG X.  
 PA (MORO/) MOROZOV I.  
 PA (PAUL/) PAUL P S.  
 PI Halbur P, Meng X, Morozov I, Paul PS;  
 DR WPI: 97-108645/10.  
 DR N-PSDB; T60795.  
 PT Porcine reproductive and respiratory syndrome virus DNA sequences -  
 PT useful for diagnosis, treatment and prevention of infection in pigs  
 PS Disclosure: Fig 2d; 114pp; English.  
 CC W25950-W25977 represent proteins encoded by ORFs 2-5 of different  
 CC isolates of porcine reproductive and respiratory syndrome virus (PRRSV).  
 CC PRRSV is a new and severe disease in swine, characterised by reproductive  
 CC failure in sows and gilts, pneumonia in young growing pigs, and an  
 CC increase in preweaning mortality. However, there are marked differences in  
 CC pathogenicity between isolates (with ISU3927 being the least virulent  
 CC isolate known). The genomic organisation of PRRSV resembles coronaviruses  
 CC and toroviruses, in that their replication involves the formation of a  
 CC 3'-coterminal nested set of subgenomic mRNAs. ORFs 5, 6, and 7 encode a  
 CC glycosylated membrane protein, an unglycosylated membrane protein, and a

CC nucleocapsid protein, respectively. ORFs 2 to 4 encode proteins with the  
 CC characteristics of membrane associated proteins. The polynucleotides of  
 CC the invention, encode a protein that is at least 88%, but less than 100%  
 CC homologous to one of proteins encoded by one of the ORFs of these  
 CC sequences. The polynucleotides of the invention, and their encoded  
 CC polypeptides can be used in a vaccine to protect a pig against PRRSV.  
 CC Antibodies raised against the polypeptides can be used to treat a pig  
 CC suffering from PRRSV, and to assay for a PRRSV.  
 SQ Sequence 200 AA;  
 Query Match 59.6%; Score 53; DB 1; Length 200;  
 Best Local Similarity 54.5%; Pred. No. 7.99e+01;  
 Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;  
 Db 19 CIVPSCFVALV 29  
 QY 1 CILESCFRAVI 11  
 ||: ||| |::  
 RESULT 11  
 ID R28757 standard; Protein; 455 AA.  
 AC R28757;  
 DT 14-JAN-1993 (first entry)  
 DE Hepatocyte nuclear factor 4. (HNF4)  
 KW hepatocyte nuclear factor 4; apolipoproteins; fat; cholesterol;  
 KW Apo CIII; Apo AI; Apo B; pyruvate kinase; alpha 1 antitrypsin;  
 KW glutamine synthetase; coronary heart disease hyperlipidaemia;  
 KW liver disease; arteriosclerosis; obesity; ss.  
 PN W09211365-A.  
 PD 09-JUL-1992.  
 PR 23-DEC-1991; U09733.  
 PR 21-DEC-1990; US-631720.  
 PA (UVRQ) UNIV ROCKEFELLER.  
 PI Darnell JE, Sladek FM, Zhong W;  
 DR WPI: 92-250087/30.  
 DR N-PSDB; Q31765.  
 PT Hepatocyte nuclear factor 4 and its DNA, regulation and  
 PT antibodies - useful for treating cardiovascular diseases e.g.  
 PT arteriosclerotic heart disease, hyperlipidaemia and  
 PT arteriosclerosis; also as an anorectic  
 PS Disclosure: Fig 3; 100pp; English.  
 CC This sequence was deduced from the cDNA sequence. The protein has  
 CC a structure analogous to that of steroid/hormone receptors. It  
 CC contains a region with two potential zinc fingers between amino  
 CC acids 50 and 116, which is 40 to 63% identical to the zinc finger  
 CC (DNA binding) domain of other members of the steroid receptor  
 CC superfamily. The proposed regulatory protein for mouse MHC I  
 CC (H2-RiBP) had the greatest similarity (62.7%), with human thyroid  
 CC hormone receptor (c-erbA, T3-R8) having 59.7% identity in this  
 CC region. The zinc finger domain is flanked by regions with no known  
 CC similarity, but there is a large hydrophobic region in the C  
 CC terminal half (133-373) which has definite similarity to the ligand  
 CC binding domain of other receptors (20-37% identity), with H-2RIIBP  
 CC being most similar at 37.3% identity. The protein also has a  
 CC proline rich region (23%) at the C terminus (400-477) which could  
 CC be an activator domain, and three serine/threonine rich regions  
 CC (30-38%) scattered through the molecule which may be  
 CC phosphorylation sites. It is not known if HNF-4 is modified,  
 CC but some post translational modification is suggested by a  
 CC molecular weight of 54KD by SDS-PAGE, but 50.8KD from the predicted  
 CC amino acid sequence. The protein itself, the gene encoding it,  
 CC Abs, and antidiabetic Abs may be used to develop diagnostic and  
 CC therapeutic agents to detect, inhibit or enhance binding to HNF-4  
 CC They can be used to study, diagnose, prevent and treat diseases  
 CC such as coronary heart disease, hyperlipidaemia, liver disease and  
 CC arteriosclerosis. They may also be used in the treatment of  
 CC obesity.  
 SQ Sequence 455 AA;  
 Query Match 59.6%; Score 53; DB 1; Length 455;  
 Best Local Similarity 66.7%; Pred. No. 7.99e+01;  
 Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

CC (Revised entry submitted to correct crossreference to N-PSDB.)  
SQ Sequence 477 AA;

Query Match 62.9%; Score 56; DB 1; Length 477;  
Best Local Similarity 40.0%; Pred. No. 3.83e+01;  
Matches 4; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Db 396 CUIQCEFRVL 405

1:::1111;

QY 1 CILESCFRAV 10

RESULT 6

ID W72387 standard; Protein; 101 AA.

AC W72387;

DT 02-FEB-1999 (first entry)

DE Pathogen response protein LSD1-interacting protein GG.

KW LSD1-interacting protein GG; plant pathogen response; apoptosis;

KW programmed cell death; disease resistance; herbicide resistance;

KW transgenic plant; crop protection.

OS Arabidopsis thaliana.

PN WO9837755-A1.

PD 03-SEP-1998.

PF 27-FEB-1998; U04077.

PR 28-FEB-1997; US-039063.

PA (UNCL-) UNIV NORTH CAROLINA.

PI Daugl JL, Dietrich RA, Epple PM, Richberg MH;

DR WPI: 98-531501/45.

DR N-PSDB; V66758.

PT New isolated Arabidopsis genes - useful for producing transgenic  
PT plants which show resistance to cell death caused by pathogens or  
PT herbicides.

PS Claim 46; Page 57; 88pp; English.

CC This is the amino acid sequence of LSD1-interacting protein GG of  
CC Arabidopsis thaliana. LSD1 interacting genes (see V66755-67) were

CC isolated from a yeast gene expression library constructed in

CC plasmid pJG4-5 using RNA from Arabidopsis leaves infected with

CC Pseudomonas syringae. A two-hybrid system was used with LSD1 short

CC and long open reading frames (see V66750-51) as bait. LSD1 (see

CC W72366-67) is a novel polypeptide that regulates the initial

CC response of plants to pathogens and the subsequent spread of plant

CC cell death engendered by infection. Since the inactivation of

CC LSD1 by mutation leads to enhanced disease resistance, LSD1

CC partner proteins represent novel targets for engineering plants

CC with enhanced resistance to pathogens. Thus, the invention

CC includes all proteins (see W72384-96) that interact with the cell

CC death regulator LSD1.

SQ Sequence 101 AA;

Query Match 60.7%; Score 54; DB 1; Length 101;

Best Local Similarity 57.1%; Pred. No. 6.26e+01;

Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 95 CLLDACF 101

1:::111

QY 1 CILESCF 7

RESULT 7

ID R73821 standard; peptide; 9 AA.

AC R73821;

DT 22-JUN-1995 (first entry)

DE Antigen fragment 137, from MAGE1 has binding affinity for HLA-2.1.

KW antigen; epitope; immunogenic target protein; PSA; HBVC; EBV;

KW HIV1; plasma specific antigen; hepatitis B virus; Epstein Barr;

KW human immunodeficiency virus; human papilloma virus; p53; c-ERB2;

KW MAGE-1; melanoma antigen-1; core antigen; surface antigen;

KW pharmaceutical composition; in vivo; ex vivo; therapeutic;

KW diagnostic; MHC class I molecule; major histocompatibility complex;

KW HLA-A2.1; 9mer; 10mer; anchor; human leukocyte antigen; PLP; 8mer;

KW algorithm prediction; MBP; CMV; cytomegalovirus; HSV;

KW herpes simplex virus; influenza A; M1; LCMV.

OS Homo sapiens.

PN WO9420127-A.  
PD 15-SEP-1994.  
PF 04-MAR-1994; U02353.  
PR 05-MAR-1993; US-027146.  
PR 04-JUN-1993; US-073205.  
PR 29-NOV-1993; US-159184.  
PA (CYTE-) CYTEL CORP.  
PI Grey HM, Kast WM, Sette A, Sidney J;  
DR WPI: 94-302678/37.  
PT Immunogenic peptide(s) having an HLA-A2.1 binding motif - used  
PT for treatment or prophylaxis of cancer, virus infection or  
PT autoimmune diseases.  
PS Disclosure; Page 85; 138pp; English.  
CC R73685-876 are potential peptide binders of HLA-A2.1 motif. Using  
CC motifs disclosed in the invention, these peptides were screened for  
CC further motifs. Only peptides with binding affinity of at least 1%  
CC (binding affinity is expressed as an IC50 value) as compared to the  
CC standard peptide (R71293) in assays. This peptide from MAGE1 has a  
CC binding value of 0.0460. The peptides of the invention can induce  
CC cytotoxic T lymphocytes which can react with target cells. They can  
CC be used for the treatment or prophylaxis of cancer, eg. prostate  
CC cancer or lymphoma, etc.  
SQ Sequence 9 AA;

Query Match 59.6%; Score 53; DB 1; Length 9;  
Best Local Similarity 88.9%; Pred. No. 7.99e+01;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 1 CILESLFRA 9

1111111111

QY 1 CILESCFRA 9

RESULT 8

ID R78910 standard; peptide; 9 AA.

AC R78910;

DT 27-MAR-1996 (first entry)

DE MAGE 1 92-100 cytotoxic T lymphocyte epitope.

KW MAGE 1 92-100; cytotoxic T; CTL; epitope; helper T; HTL; cell;

KW lymphocyte; antigens; treatment; disease prevention; tumours;

KW cancer; melanomas.

OS Homo sapiens.

PN WO9522317-A1.

PD 24-AUG-1995.

PF 16-FEB-1995; U02121.

PR 16-FEB-1994; US-197484.

PA (CYTE-) CYTEL CORP.

PI Ceut RW, Grey H, Sette AD, Vitello MA;

DR WPI: 95-302545/39.

PT Compn. inducing cytotoxic T lymphocyte response to pref. viral,

PT bacterial, parasitic or tumour antigens - useful in the treatment

PT and prevention of diseases associated with the antigen e.g.

PT hepatitis B

PS Example 13; Page 71; 109pp; English.

CC A compn. which induces a cytotoxic T lymphocyte (CTL) response to

CC a human MAGE antigen (Ag) in a mammal comprises, a MAGE CTL Ag

CC response inducing peptide (i.e. R78904 to R78917) and a lipid

CC conjugated helper T cell inducing peptide. The compn. is useful

CC in the treatment and prevention of MAGE tumour Ag associated

CC diseases, e.g. melanoma cancers.

SQ Sequence 9 AA;

Query Match 59.6%; Score 53; DB 1; Length 9;  
Best Local Similarity 88.9%; Pred. No. 7.99e+01;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 1 CILESLFRA 9

1111111111

QY 1 CILESCFRA 9

RESULT 9

ID R94701 standard; Protein; 200 AA.

DE Human melanoma antigen MAGE-1.  
 KW HLA-restricted cytotoxic T-lymphocyte activity.  
 OS Homo sapiens.  
 PN W09504542-A.

PD 16-FEB-1995.  
 PF 02-AUG-1994; 008721.  
 PR 06-AUG-1993; US-103623.  
 PA (CYTE-) CYTEL CORP.  
 PI Fikes JD, Livingston BD, Sette AD, Sidney JC;  
 DR WPI: 95-090681/12.  
 DR N-PSDB; Q85435.  
 PT Human melanoma antigen, MAGE-1, peptide(s) - useful for  
 stimulating immune response against melanoma  
 PS Example 1: Fig 1: 59pp; English.  
 CC Q85435 encodes R70909 human melanoma antigen MAGE-1, it was used  
 CC to produce the C-terminal MAGE-1 peptides described in R70915 to  
 CC R70969. These peptides are useful for defining epitopes that  
 CC engender a HLA-restricted cytotoxic lymphocyte activity against  
 CC MAGE-1 antigens. Comps. containing these peptides can be  
 CC administered, as a vaccine to patients susceptible to MAGE  
 CC associated tumours, e.g. melanomas.  
 SQ Sequence 309 AA;

Query Match 75.3%; Score 67; DB 1; Length 309;  
 Best Local Similarity 90.9%; Pred. No. 2.35e+00;  
 Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 92 CILESFRAVI 102

QY 1 CILESCFRAVI 11

#### RESULT 3

ID W81548 standard; Protein; 309 AA.  
 AC W81548;  
 DT 01-MAR-1999 (first entry)  
 DE Tumour rejection antigen precursor MAGE-A1.  
 KW MAGE-A1; human; tumour rejection antigen precursor; TRAP;  
 KW therapy; diagnosis.  
 OS Homo sapiens.  
 PN W09849184-A1.  
 PD 05-NOV-1998.  
 PR 24-APR-1998; US-845528.  
 PA (LUDW-) LUDWIG INST CANCER RES.  
 PI Boon-Falleur T, De Smet C, Lucas S;  
 DR WPI: 99-024041/02.  
 DR N-PSDB; V69719.  
 PT Tumour rejection antigen precursors - used for determining presence  
 PT of cytolytic T cells specific for complexes of a human leukocyte  
 PT antigen  
 PS Disclosure; Page 50-51; 84pp; English.  
 CC This is the amino acid sequence of human tumour rejection antigen  
 CC precursor (TRAP) MAGE-A1. MAGE-A1 cDNA (see V69719) shows homology  
 CC to novel human MAGE-C1 cDNA (see W81546). MAGE-C1 is a  
 CC novel member of the MAGE family that may be recognised by cytotoxic  
 CC T cells, leading to lysis of the tumour cells which express it. It  
 CC is expressed in a variety of tumours and in normal testis cells.  
 CC but not by other normal cells. The invention provides MAGE-C1 and  
 CC MAGE-C2 nucleic acids and polypeptides, useful e.g. in a claimed  
 CC method for determining the presence of cytolytic T cells specific  
 CC for complexes of a human leukocyte antigen (HLA).  
 SQ Sequence 309 AA;

Query Match 75.3%; Score 67; DB 1; Length 309;  
 Best Local Similarity 90.9%; Pred. No. 2.35e+00;  
 Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 92 CILESFRAVI 102

QY 1 CILESCFRAVI 11

#### RESULT 4

ID R31349 standard; Protein; 615 AA.  
 AC R31349;  
 DT 18-MAY-1993 (first entry)  
 DE Jaagsiekte retrovirus Env protein.  
 KW JSRV; epithelial carcinoma; ovine; sheep; vaccine;  
 KW pulmonary adenomatosa; envelope glycoprotein.  
 OS Jaagsiekte retrovirus.  
 FH Key Location/Qualifiers  
 FT region 1..378  
 FT /note= "surface portion"  
 FT /note= "transmembrane portion"  
 FT region 379..615  
 PN FR2676455-A.  
 PD 20-NOV-1992.  
 PF 17-MAY-1991; 006060.  
 PR 17-MAY-1991; FR-006060.  
 PA (INRM) INSERM INST NAT SANTE & RECH MED.  
 PI Querat GF, Verwoerd D, Vigne R, York D;  
 DR WPI: 93-020250/03.  
 DR N-PSDB; Q35153.  
 PT New Jaagsiekte Retrovirus and characteristic nucleic acid - also  
 PT derived proteins, probes and antibodies, useful for in vitro  
 PT diagnosis and in vaccines  
 PS Claim 26; Page 41-43; 75pp; French.  
 CC JSRV causes epithelial carcinoma in ovine animals, partic. pulmonary  
 CC adenomatosa in sheep. The complete cDNA sequence of the JSRV genome  
 CC was prepared from an approx. 8.7kb band of poly-A RNA isolated from  
 CC semi-purified lung lavage samples from infected sheep. The  
 CC invention includes the Env amino acid sequence or any part of it  
 CC which is capable of specific immunological reaction with antibodies  
 CC directed against JSRV. The glycoproteins gp46 or gp31 and the  
 CC precursor PR69 env are preferred  
 CC See also R31346-R31348 and Q35153-Q35155.  
 SQ Sequence 615 AA;

Query Match 65.2%; Score 58; DB 1; Length 615;  
 Best Local Similarity 60.0%; Pred. No. 2.33e+01;  
 Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 323 CILTNCIRGV 332

QY 1 CILESCFRAV 10

#### RESULT 5

ID W25153 standard; Protein; 477 AA.  
 AC W25153;  
 DT 22-JAN-1998 (updated)  
 DT 03-DEC-1997 (first entry)  
 DE Nsp7524III restriction enzyme isoform.  
 KW Restriction enzyme; NspIII; Nostoc species; genetic engineering;  
 KW cloning; vector construction; recombinant production; endonuclease.  
 OS Nostoc sp.  
 PN J09191885-A.  
 PD 29-JUL-1997.  
 PF 16-JAN-1996; 023304.  
 PR 16-JAN-1996; JP-023304.  
 PA (TAKI) TAKARA SHUZO CO LTD.  
 DR WPI: 97-429185/40.  
 DR N-PSDB; T79876, T89627.  
 PT Nsp7524III restriction endonuclease and its gene - useful in  
 PT genetic engineering methods, e.g. vector construction and cloning  
 PS Claim 9; Page 11-12; 15pp; Japanese.  
 CC W25153 shows the sequence of an Nsp7524III restriction enzyme isoform  
 CC (derived from Nostoc sp. PCC7524). The full length gene encoding this  
 CC enzyme also contains a second open reading frame encoding a similar  
 CC but different NspIII enzyme.  
 CC NspIII restriction endonucleases are useful in genetic engineering  
 CC methods such as vector construction and cloning. The enzymes cut  
 CC between the first and second nucleotides of the sequence CYCGRG.  
 CC Nsp7524III can be produced recombinantly in a large amount.

\*\*\*\*\*

W P S R L H  
 (TM)

\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
 Copyright (C) 1993-1998 University of Edinburgh, U.K.  
 Distribution rights by Oxford Molecular Ltd

MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Apr 14 23:11:51 2000; Maspar time 4.65 Seconds  
 Tabular output not generated. 56.016 Million cell updates/sec

Title: >US-08-452-843-3  
 Description: (1-11) from US08452843.pgp  
 Perfect Score: 89  
 Sequence: 1 CILESCFRAVI 11

Scoring table: PAM 150  
 Gap 15

Searched: 188963 seqs, 23686106 residues

Post-processing: Minimum Match 0%  
 Listing first 45 summaries

Database: a:geneseq36  
 1:geneseq36

Statistics: Mean 17.860; Variance 54.335; scale 0.329

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	89	100.0	11	1 R89364	MAGE-1 derived immunog	6.37e-03
2	67	75.3	309	1 R70909	Human melanoma antigen	2.35e+00
3	67	75.3	309	1 W81548	Tumour rejection antig	2.35e+00
4	58	65.2	615	1 R31349	Jaagsiekte retrovirus	2.33e+01
5	56	62.9	477	1 W25153	Nsp7524III restriction	3.83e+01
6	54	60.7	101	1 W72387	Pathogen response prot	6.26e+01
7	53	59.6	9	1 R73821	Antigen fragment 137,	7.99e+01
8	53	59.6	9	1 R78910	MAGE-1 92-100 cytotoxi	7.99e+01
9	53	59.6	200	1 R4701	PRRSV VR 2385 ORF-5 pr	7.99e+01
10	53	59.6	200	1 R25953	ORF 5 protein of PRRSV	7.99e+01
11	53	59.6	455	1 R28757	Hepatocyte nuclear fac	7.99e+01
12	53	59.6	465	1 W71574	Human native hepatocyt	7.99e+01
13	53	59.6	516	1 W71587	Human hepatocyte nucle	7.99e+01
14	53	59.6	567	1 W71582	Human hepatocyte nucle	7.99e+01
15	52	58.4	489	1 R42895	Corticosterone-binding	1.02e+02
16	51	57.3	328	1 W87723	H. pylori GHPO 1196 pr	1.30e+02
17	50	56.2	308	1 R87723	Full length coconut LP	1.64e+02
18	50	56.2	358	1 W20718	H. pylori membrane pro	1.64e+02
19	50	56.2	591	1 R74802	Saccharomyces sp. reco	1.64e+02
20	50	56.2	832	1 W74089	Human HPT-1 protein se	1.64e+02
21	49	55.1	34	1 W88665	Secreted protein encod	2.09e+02
22	49	55.1	35	1 W82286	Duodenal lumen to brai	2.09e+02
23	49	55.1	251	1 W47421	Bacillus subtilis pren	2.09e+02

24	49	55.1	286	1 W20102	H. pylori cytoplasmic	2.09e+02
25	49	55.1	286	1 W24585	H. pylori cytoplasmic	2.09e+02
26	49	55.1	455	1 W20606	H. pylori cytoplasmic	2.09e+02
27	49	55.1	500	1 W30843	Partial rat thrombomod	2.09e+02
28	49	55.1	559	1 W30844	Partial rat thrombomod	2.09e+02
29	49	55.1	577	1 W30845	Rat thrombomodulin.	2.09e+02
30	49	55.1	889	1 R56248	Xenopus thrombospondin	2.09e+02
31	49	55.1	1261	1 W75995	GPase activating prot	2.09e+02
32	49	55.1	3164	1 R94345	Hepatitis GB virus (HG	2.09e+02
33	48	53.9	55	1 R63334	HT-LCF fragment, corre	2.64e+02
34	48	53.9	94	1 R63332	Leukocyte Chemotactic	2.64e+02
35	48	53.9	106	1 R63353	Recombinant (Met)-HT-L	2.64e+02
36	48	53.9	337	1 W40137	Human partial GalR2 re	2.64e+02
37	48	53.9	385	1 W52352	Human galanin receptor	2.64e+02
38	48	53.9	387	1 W61386	Human galanin receptor	2.64e+02
39	48	53.9	387	1 W49003	Human galanin receptor	2.64e+02
40	48	53.9	387	1 W24562	Human galanin receptor	2.64e+02
41	48	53.9	498	1 W53461	Human latheo protein 1	2.64e+02
42	48	53.9	556	1 W53459	Human latheo protein s	2.64e+02
43	48	53.9	580	1 R93607	Kaposi's sarcoma assoc	2.64e+02
44	48	53.9	580	1 R97831	Kaposi's sarcoma assoc	2.64e+02
45	48	53.9	896	1 R63333	Human HT-1376 cell-der	2.64e+02

## ALIGNMENTS

RESULT 1  
 ID R89364 standard; peptide; 11 AA.  
 AC R89364; 1996 (first entry)  
 DT 18-SEP-1996  
 DE MAGE-1 derived immunogenic peptide.  
 KW Immunogenic peptide; supermotif; HLA molecule; CTL response;  
 KW therapeutic; diagnostic; cancer; viral infection; hepatitis B;  
 KW hepatitis C.  
 OS Synthetic.  
 PN WO9603140-A1.  
 PD 08-FEB-1996.  
 PF 21-JUL-1995; U09234.  
 PR 21-JUL-1994; US-278634.  
 PR 23-NOV-1994; US-344824.  
 PR 30-MAY-1995; US-452843.  
 PA (CYTE-) CYTEL CORP.  
 PI Sette A, Sidney J;  
 DT WPI; 96-116784/12.  
 PT Compsn. comprising immunogenic peptide with supermotif allowing more  
 PT than one HLA mol. to bind - used to induce CTL response in patient  
 PT and for in vivo and ex vivo therapeutic and diagnostic applications  
 PS Claim 2; Page 26; 32pp; English.  
 CC The sequences given in R89362-82 are immunogenic peptides which were  
 CC use in the composition of the invention. The composition comprises  
 CC an immunogenic peptide of 9-10 residues with a supermotif which  
 CC allows binding of more than one HLA molecule. It pref. comprises  
 CC two conserved residues, a first at the 2nd position from the N-  
 CC terminal is pro, and a 2nd at the C-terminal is Met. These peptides  
 CC are used to induce a CTL response in a patient. They are also  
 CC useful in compositions for in vivo and ex vivo therapeutic and  
 CC diagnostic applications, e.g the treatment of cancer and viral  
 CC infections, e.g. hepatitis B and C.  
 SQ Sequence 11 AA;

Query Match 100.0%; Score 89; DB 1; Length 11;  
 Best Local Similarity 100.0%; Pred. No. 6.37e-03;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 CILESCFRAVI 11

QY 1 CILESCFRAVI 11

RESULT 2

ID R70909 standard; Protein; 309 AA.  
 AC R70909;  
 DT 09-OCT-1995 (first entry)



THIS PAGE BLANK (USPTO)

DT 01-NOV-1999 (TReMBLrel. 12, Last sequence update)  
 DT 01-NOV-1999 (TReMBLrel. 12, Last annotation update)  
 DE REVERSE TRANSCRIPTASE (FRAGMENT).  
 GN POL.  
 OS Human immunodeficiency virus type 1.  
 OC Viruses; Retrovirdae; Retroviridae; Lentivirus.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 97281566.  
 RA NAKANO T., MOROZUMI H., INUZUKA S., NAGATA M., TAGUCHI Y.,  
 RA MIZOKAMI M., OKAMOTO T.,  
 RT "Clonal selection of HIV type 1 variants associated with resistance to  
 RL foscarnet in vitro: confirmation by molecular evolutionary analysis.";  
 RL AIDS Res. Hum. Retroviruses 13:563-573(1997).  
 DR EMBL: D78529; BAA24331.1; -.  
 DR EMBL: D78516; BAA24272.1; -.  
 DR EMBL: D78527; BAA24329.1; -.  
 KW RNA-directed DNA polymerase.  
 FT NON\_TER 1  
 FT NON\_TER 263  
 SQ SEQUENCE 263 AA; 30670 MW; E79E4D69 CRC32;  
  
 Query Match 85.4%; Score 70; DB 14; Length 263;  
 Best Local Similarity 100.0%; Pred. No. 3.05e-03;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
 Db 11 PKVKQWPL 18  
 QY 2 PKVKQWPL 9  
 |||||  
 |||||

Search completed: Fri Apr 14 23:09:54 2000  
 Job time : 98 secs.

```
Db 17 PKVKQWPL 24
QY 2 PKVKQWPL 9

RESULT 11
ID Q9YU08 PRELIMINARY; PRT; 218 AA.
AC Q9YU08
DT 01-MAY-1999 (TRENBLrel. 10, Created)
DT 01-MAY-1999 (TRENBLrel. 10, Last sequence update)
DT 01-MAY-1999 (TRENBLrel. 10, Last annotation update)
DE REVERSE TRANSCRIPTASE (FRAGMENT).
GN POL.
OS Human immunodeficiency virus type 1.
OC Viruses; Retrovirdae; Retroviridae; Lentivirus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-PATIENT 22;
RX MEDLINE; 99019109.
RA ORLANDI P., CANCRINI C., SCACCIA S., ROMITI M.L., LIVADIOTTI S.,
RA GATTINARA G.C., ANGELINI F., COX S., ROSSI P.;
RT "Analysis of HIV-1 reverse transcriptase gene mutations in infected
RT children treated with zidovudine."
RL J. Acquir. Immune Defic. Syndr. Hum. Retrovirol. 19:230-237(1998).
DR EMBL; AF056681; RAD19271.1; -.
DR EMBL; AF056680; RAD19270.1; -.
KW RNA-directed DNA polymerase.
FT NON_TER 1
FT NON_TER 218
SQ SEQUENCE 218 AA; 25291 MW; AA688AB9 CRC32;

Query Match 85.4%; Score 70; DB 14; Length 218;
Best Local Similarity 100.0%; Pred. No. 3.05e-03;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 17 PKVKQWPL 24
QY 2 PKVKQWPL 9

RESULT 12
ID Q9WH1 PRELIMINARY; PRT; 218 AA.
AC Q9WH1
DT 01-NOV-1999 (TRENBLrel. 12, Created)
DT 01-NOV-1999 (TRENBLrel. 12, Last sequence update)
DT 01-NOV-1999 (TRENBLrel. 12, Last annotation update)
DE REVERSE TRANSCRIPTASE (FRAGMENT).
GN POL.
OS Human immunodeficiency virus type 1.
OC Viruses; Retrovirdae; Retroviridae; Lentivirus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-PATIENT 15;
RX MEDLINE; 99019109.
RA ORLANDI P., CANCRINI C., SCACCIA S., ROMITI M.L., LIVADIOTTI S.,
RA GATTINARA G.C., ANGELINI F., COX S., ROSSI P.;
RT "Analysis of HIV-1 reverse transcriptase gene mutations in infected
RT children treated with zidovudine."
RL J. Acquir. Immune Defic. Syndr. Hum. Retrovirol. 19:230-237(1998).
DR EMBL; AF056681; RAD19258.1; -.
DR EMBL; AF056680; RAD19257.1; -.
KW RNA-directed DNA polymerase.
FT NON_TER 1
FT NON_TER 218
SQ SEQUENCE 218 AA; 25193 MW; AA00C86E CRC32;

Query Match 85.4%; Score 70; DB 14; Length 218;
Best Local Similarity 100.0%; Pred. No. 3.05e-03;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 17 PKVKQWPL 24
QY 2 PKVKQWPL 9

RESULT 13
ID Q9YQS3 PRELIMINARY; PRT; 259 AA.
AC Q9YQS3
DT 01-MAY-1999 (TRENBLrel. 10, Created)
DT 01-MAY-1999 (TRENBLrel. 10, Last sequence update)
DT 01-NOV-1999 (TRENBLrel. 12, Last annotation update)
DE REVERSE TRANSCRIPTASE (FRAGMENT).
OS Human immunodeficiency virus type 1.
OC Viruses; Retrovirdae; Retroviridae; Lentivirus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-L2428;
RA SCHMIT J.C.C., RUIZ L., HERMANS P., SOENNERBORG A., LEAL M.,
RA HARRER T., CLOTET B., SPRECHER S., ARENDT V., LISSEN E., WIIVROUT M.,
RA DESMYTER J., DE CLERCQ E., VANDAMME A.M.;
RT "Multiple dideoxynucleoside analogue-resistant (MddNR) HIV-1 strains
RT isolated from patients in different European countries."
RL AIDS 12:2005-2015(1998).
DR EMBL; AJ003203; CAA05983.1; -.
DR HSSP; P03366; IDLO.
KW RNA-directed DNA polymerase.
FT NON_TER 1
FT NON_TER 259
SQ SEQUENCE 259 AA; 30074 MW; A5001EA3 CRC32;

Query Match 85.4%; Score 70; DB 14; Length 259;
Best Local Similarity 100.0%; Pred. No. 3.05e-03;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 19 PKVKQWPL 26
QY 2 PKVKQWPL 9

RESULT 14
ID Q9YQS1 PRELIMINARY; PRT; 259 AA.
AC Q9YQS1
DT 01-MAY-1999 (TRENBLrel. 10, Created)
DT 01-MAY-1999 (TRENBLrel. 10, Last sequence update)
DT 01-NOV-1999 (TRENBLrel. 12, Last annotation update)
DE REVERSE TRANSCRIPTASE (FRAGMENT).
OS Human immunodeficiency virus type 1.
OC Viruses; Retrovirdae; Retroviridae; Lentivirus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-DE296;
RA SCHMIT J.C.C., RUIZ L., HERMANS P., SOENNERBORG A., LEAL M.,
RA HARRER T., CLOTET B., SPRECHER S., ARENDT V., LISSEN E., WIIVROUT M.,
RA DESMYTER J., DE CLERCQ E., VANDAMME A.M.;
RT "Multiple dideoxynucleoside analogue-resistant (MddNR) HIV-1 strains
RT isolated from patients in different European countries."
RL AIDS 12:2005-2015(1998).
DR EMBL; AJ003212; CAA05992.1; -.
DR HSSP; P03366; IDLO.
KW RNA-directed DNA polymerase.
FT NON_TER 1
FT NON_TER 259
SQ SEQUENCE 259 AA; 30125 MW; BA43AEF9 CRC32;

Query Match 85.4%; Score 70; DB 14; Length 259;
Best Local Similarity 100.0%; Pred. No. 3.05e-03;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 19 PKVKQWPL 26
QY 2 PKVKQWPL 9

RESULT 15
ID Q9W8A4 PRELIMINARY; PRT; 263 AA.
AC Q9W8A4
DT 01-NOV-1999 (TRENBLrel. 12, Created)
```

KW RNA-directed DNA polymerase.

FT NON\_TER 1  
SQ SEQUENCE 218 AA; 25269 MW; B3C280B7 CRC32;

Query Match 85.4%; Score 70; DB 14; Length 218;  
Best Local Similarity 100.0%; Pred. No. 3.05e-03;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 17 PKVKQWPL 24  
QY 2 PKVKQWPL 9

RESULT 7  
ID Q9W8M2 PRELIMINARY; PRT; 218 AA.  
AC Q9W8M2;  
DT 01-NOV-1999 (Tremblrel. 12, Created)  
DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)  
DT 01-NOV-1999 (Tremblrel. 12, Last annotation update)  
DE REVERSE TRANSCRIPTASE (FRAGMENT).  
GN POL.  
OS Human immunodeficiency virus type 1.  
OC Viruses; Retrov. Retroviridae; Lentivirus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-PATIENT 2;  
RX MEDLINE; 99019109.  
RA ORLANDI P., CANCRINI C., SCACCIA S., ROMITI M.L., LIVADIOTTI S.,  
RA GATTINARA G.C., ANGELINI F., COX S., ROSSI P.;  
RT "Analysis of HIV-1 reverse transcriptase gene mutations in infected  
RT children treated with zidovudine."  
RL J. Acquir. Immune Defic. Syndr. Hum. Retrovirol. 19:230-237(1998).  
DR EMBL; AF056641; AAD19231.1; -  
DR EMBL; AF056640; AAD19230.1; -  
KW RNA-directed DNA polymerase.  
FT NON\_TER 1  
FT NON\_TER 218  
SQ SEQUENCE 218 AA; 25215 MW; BC4A2548 CRC32;

Query Match 85.4%; Score 70; DB 14; Length 218;  
Best Local Similarity 100.0%; Pred. No. 3.05e-03;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 17 PKVKQWPL 24  
QY 2 PKVKQWPL 9

RESULT 8  
ID Q9W8G3 PRELIMINARY; PRT; 218 AA.  
AC Q9W8G3;  
DT 01-NOV-1999 (Tremblrel. 12, Created)  
DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)  
DT 01-NOV-1999 (Tremblrel. 12, Last annotation update)  
DE REVERSE TRANSCRIPTASE (FRAGMENT).  
GN POL.  
OS Human immunodeficiency virus type 1.  
OC Viruses; Retrov. Retroviridae; Lentivirus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-PATIENT 4;  
RX MEDLINE; 99019109.  
RA ORLANDI P., CANCRINI C., SCACCIA S., ROMITI M.L., LIVADIOTTI S.,  
RA GATTINARA G.C., ANGELINI F., COX S., ROSSI P.;  
RT "Analysis of HIV-1 reverse transcriptase gene mutations in infected  
RT children treated with zidovudine."  
RL J. Acquir. Immune Defic. Syndr. Hum. Retrovirol. 19:230-237(1998).  
DR EMBL; AF056644; AAD19234.1; -  
KW RNA-directed DNA polymerase.  
FT NON\_TER 1  
FT NON\_TER 218  
SQ SEQUENCE 218 AA; 25241 MW; 15A39EBA CRC32;

Query Match 85.4%; Score 70; DB 14; Length 218;  
Best Local Similarity 100.0%; Pred. No. 3.05e-03;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 17 PKVKQWPL 24  
QY 2 PKVKQWPL 9

RESULT 9  
ID Q9WBG2 PRELIMINARY; PRT; 218 AA.  
AC Q9WBG2;  
DT 01-NOV-1999 (Tremblrel. 12, Created)  
DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)  
DT 01-NOV-1999 (Tremblrel. 12, Last annotation update)  
DE REVERSE TRANSCRIPTASE (FRAGMENT).  
GN POL.  
OS Human immunodeficiency virus type 1.  
OC Viruses; Retrov. Retroviridae; Lentivirus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-PATIENT 3;  
RX MEDLINE; 99019109.  
RA ORLANDI P., CANCRINI C., SCACCIA S., ROMITI M.L., LIVADIOTTI S.,  
RA GATTINARA G.C., ANGELINI F., COX S., ROSSI P.;  
RT "Analysis of HIV-1 reverse transcriptase gene mutations in infected  
RT children treated with zidovudine."  
RL J. Acquir. Immune Defic. Syndr. Hum. Retrovirol. 19:230-237(1998).  
DR EMBL; AF056642; AAD19232.1; -  
KW RNA-directed DNA polymerase.  
FT NON\_TER 1  
FT NON\_TER 218  
SQ SEQUENCE 218 AA; 25144 MW; 92B78FD2 CRC32;

Query Match 85.4%; Score 70; DB 14; Length 218;  
Best Local Similarity 100.0%; Pred. No. 3.05e-03;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 17 PKVKQWPL 24  
QY 2 PKVKQWPL 9

RESULT 10  
ID Q9W9R7 PRELIMINARY; PRT; 218 AA.  
AC Q9W9R7;  
DT 01-NOV-1999 (Tremblrel. 12, Created)  
DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)  
DT 01-NOV-1999 (Tremblrel. 12, Last annotation update)  
DE REVERSE TRANSCRIPTASE (FRAGMENT).  
GN POL.  
OS Human immunodeficiency virus type 1.  
OC Viruses; Retrov. Retroviridae; Lentivirus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-PATIENT 17;  
RX MEDLINE; 99019109.  
RA ORLANDI P., CANCRINI C., SCACCIA S., ROMITI M.L., LIVADIOTTI S.,  
RA GATTINARA G.C., ANGELINI F., COX S., ROSSI P.;  
RT "Analysis of HIV-1 reverse transcriptase gene mutations in infected  
RT children treated with zidovudine."  
RL J. Acquir. Immune Defic. Syndr. Hum. Retrovirol. 19:230-237(1998).  
DR EMBL; AF056671; AAD19261.1; -  
DR EMBL; AF056670; AAD19260.1; -  
KW RNA-directed DNA polymerase.  
FT NON\_TER 1  
FT NON\_TER 218  
SQ SEQUENCE 218 AA; 25170 MW; 30759373 CRC32;

Query Match 85.4%; Score 70; DB 14; Length 218;  
Best Local Similarity 100.0%; Pred. No. 3.05e-03;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

RA SEQUENCE FROM N.A.
RC STRAIN=ZH106 FROM AUSTRALIA;
RA ZHENG N.N., HURREN L., NEILAN B.A., COOPER D.A., DELANEY S.F.,
RA MCQUEEN P.W.;
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL: U64189; AAB05332.1; -.
DR HSSP: P04585; IRT2.
DR PFAM: PF00078; rvt; 1.
KW RNA-directed DNA polymerase.
FT NON_TER 1
FT NON_TER 245
SQ SEQUENCE 245 AA; 28584 MW; 16A34C26 CRC32;

Query Match 91.5%; Score 75; DB 14; Length 245;
Best Local Similarity 88.9%; Pred. No. 2.23e-04;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 10 YPKVKQWPF 18
QY 1 YPKVKQWPL 9

RESULT 3
ID Q9W9N4 PRELIMINARY; PRT; 218 AA.
AC Q9W9N4;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
DE REVERSE TRANSCRIPTASE (FRAGMENT).
GN POL.
OS Human immunodeficiency virus type 1.
OC Viruses; Retroviral viruses; Retroviridae; Lentivirus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-PATIENT 21;
RX MEDLINE; 99019109.
RA ORLANDI P., CANCRINI C., SCACCIA S., ROMITI M.L., LIVADIOTTI S.,
RA GATTINARA G.C., ANGELINI F., COX S., ROSSI P.;
RT "Analysis of HIV-1 reverse transcriptase gene mutations in infected
RT children treated with zidovudine.";
RL J. Acquir. Immune Defic. Syndr. Hum. Retrovirol. 19:230-237(1998).
DR EMBL: AF056679; AAD19269.1; -.
DR EMBL: AF056677; AAD19267.1; -.
DR EMBL: AF056678; AAD19268.1; -.
KW RNA-directed DNA polymerase.
FT NON_TER 1
FT NON_TER 218
SQ SEQUENCE 218 AA; 25218 MW; 3E351F11 CRC32;

Query Match 85.4%; Score 70; DB 14; Length 218;
Best Local Similarity 100.0%; Pred. No. 3.05e-03;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 17 PKVKQWPL 24
QY 2 PKVKQWPL 9

RESULT 4
ID Q9WBG9 PRELIMINARY; PRT; 218 AA.
AC Q9WBG9;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
DE REVERSE TRANSCRIPTASE (FRAGMENT).
GN POL.
OS Human immunodeficiency virus type 1.
OC Viruses; Retroviral viruses; Retroviridae; Lentivirus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-PATIENT 7;
RX MEDLINE; 99019109.
RA ORLANDI P., CANCRINI C., SCACCIA S., ROMITI M.L., LIVADIOTTI S.,
RA GATTINARA G.C., ANGELINI F., COX S., ROSSI P.;
RT "Analysis of HIV-1 reverse transcriptase gene mutations in infected
RT children treated with zidovudine.";
RL J. Acquir. Immune Defic. Syndr. Hum. Retrovirol. 19:230-237(1998).
DR EMBL: AF056647; AAD19237.1; -.
DR EMBL: AF056647; AAD19237.1; -.
KW RNA-directed DNA polymerase.
FT NON_TER 1
FT NON_TER 218
SQ SEQUENCE 218 AA; 25292 MW; 268B01B8 CRC32;

Query Match 85.4%; Score 70; DB 14; Length 218;
Best Local Similarity 100.0%; Pred. No. 3.05e-03;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 17 PKVKQWPL 24
QY 2 PKVKQWPL 9

RESULT 5
ID Q9WBG5 PRELIMINARY; PRT; 218 AA.
AC Q9WBG5;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
DE REVERSE TRANSCRIPTASE (FRAGMENT).
GN POL.
OS Human immunodeficiency virus type 1.
OC Viruses; Retroviral viruses; Retroviridae; Lentivirus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-PATIENT 4;
RX MEDLINE; 99019109.
RA ORLANDI P., CANCRINI C., SCACCIA S., ROMITI M.L., LIVADIOTTI S.,
RA GATTINARA G.C., ANGELINI F., COX S., ROSSI P.;
RT "Analysis of HIV-1 reverse transcriptase gene mutations in infected
RT children treated with zidovudine.";
RL J. Acquir. Immune Defic. Syndr. Hum. Retrovirol. 19:230-237(1998).
DR EMBL: AF056647; AAD19237.1; -.
DR EMBL: AF056647; AAD19237.1; -.
KW RNA-directed DNA polymerase.
FT NON_TER 1
FT NON_TER 218
SQ SEQUENCE 218 AA; 25292 MW; 268B01B8 CRC32;

Query Match 85.4%; Score 70; DB 14; Length 218;
Best Local Similarity 100.0%; Pred. No. 3.05e-03;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 17 PKVKQWPL 24
QY 2 PKVKQWPL 9

RESULT 6
ID Q9WBG4 PRELIMINARY; PRT; 218 AA.
AC Q9WBG4;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
DE REVERSE TRANSCRIPTASE (FRAGMENT).
GN POL.
OS Human immunodeficiency virus type 1.
OC Viruses; Retroviral viruses; Retroviridae; Lentivirus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-PATIENT 4;
RX MEDLINE; 99019109.
RA ORLANDI P., CANCRINI C., SCACCIA S., ROMITI M.L., LIVADIOTTI S.,
RA GATTINARA G.C., ANGELINI F., COX S., ROSSI P.;
RT "Analysis of HIV-1 reverse transcriptase gene mutations in infected
RT children treated with zidovudine.";
RL J. Acquir. Immune Defic. Syndr. Hum. Retrovirol. 19:230-237(1998).
DR EMBL: AF056645; AAD19235.1; -.
DR EMBL: AF056645; AAD19235.1; -.

```

\*\*\*\*\*  
MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm  
\*\*\*\*\*  
Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd  
(TM)  
\*\*\*\*\*

Run on: Fri Apr 14 23:08:16 2000; Maspar.time 12.23 Seconds  
Tabular output not generated. 51.037 Million cell updates/sec.

Title: >US-08-452-843-2  
Description: (1-9) from US08452843.pep  
Perfect Score: 82  
Sequence: 1 YPKVKQWPL 9

Scoring table: PAM 150  
Gap 15

Searched: 225878 seqs, 69334122 residues  
Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: sptrembl12  
1:sp\_archaea 2:sp\_bacteria 3:sp\_fungi 4:sp\_human  
5:sp\_invertebrate 6:sp\_mammal 7:sp\_mhc 8:sp\_organelle  
9:sp\_phase 10:sp\_plant 11:sp\_rodent 12:sp\_unclassified  
13:sp\_vertebrate 14:sp\_virus

Statistics: Mean 24.437; Variance 32.175; scale 0.760  
Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES					Pred. No.	
Result No.	Score	Query Match	Length	Description	ID	Pred. No.
1	82	100.0	245 14	REVERSE TRANSCRIPTASE	Q75826	5.06e-06
2	75	91.5	245 14	REVERSE TRANSCRIPTASE	Q75832	2.23e-04
3	70	85.4	218 14	REVERSE TRANSCRIPTASE	Q9W9N4	3.05e-03
4	70	85.4	218 14	REVERSE TRANSCRIPTASE	Q9WBG9	3.05e-03
5	70	85.4	218 14	REVERSE TRANSCRIPTASE	Q9WBG5	3.05e-03
6	70	85.4	218 14	REVERSE TRANSCRIPTASE	Q9WBG4	3.05e-03
7	70	85.4	218 14	REVERSE TRANSCRIPTASE	Q9W8M2	3.05e-03
8	70	85.4	218 14	REVERSE TRANSCRIPTASE	Q9WBG3	3.05e-03
9	70	85.4	218 14	REVERSE TRANSCRIPTASE	Q9WBG2	3.05e-03
10	70	85.4	218 14	REVERSE TRANSCRIPTASE	Q9W9R7	3.05e-03
11	70	85.4	218 14	REVERSE TRANSCRIPTASE	Q9YIU8	3.05e-03
12	70	85.4	218 14	REVERSE TRANSCRIPTASE	Q9WBH1	3.05e-03
13	70	85.4	259 14	REVERSE TRANSCRIPTASE	Q9YQS3	3.05e-03
14	70	85.4	259 14	REVERSE TRANSCRIPTASE	Q9YQS1	3.05e-03
15	70	85.4	263 14	REVERSE TRANSCRIPTASE	Q9W8A4	3.05e-03
16	70	85.4	341 14	POL PROTEIN (FRAGMENT)	Q9W8A6	3.05e-03
17	70	85.4	341 14	POL PROTEIN (FRAGMENT)	Q9W8V4	3.05e-03
18	70	85.4	341 14	POL POLYPROTEIN (FRAGM	Q9WJK2	3.05e-03
19	70	85.4	341 14	POL PROTEIN (FRAGMENT)	Q9WE28	3.05e-03
20	70	85.4	341 14	POL PROTEIN (FRAGMENT)	Q9WF00	3.05e-03

21	70	85.4	341 14	Q9WF01	POL PROTEIN (FRAGMENT)	3.05e-03
22	70	85.4	341 14	Q9WF11	POL PROTEIN (FRAGMENT)	3.05e-03
23	70	85.4	347 14	Q9WFJ2	POL POLYPROTEIN (FRAGM	3.05e-03
24	70	85.4	347 14	Q9YLP6	POL POLYPROTEIN (FRAGM	3.05e-03
25	70	85.4	347 14	Q9YLP5	POL POLYPROTEIN (FRAGM	3.05e-03
26	70	85.4	347 14	Q9YLP3	POL POLYPROTEIN (FRAGM	3.05e-03
27	70	85.4	347 14	Q9WFJ6	POL POLYPROTEIN (FRAGM	3.05e-03
28	70	85.4	347 14	Q9WFJ0	POL POLYPROTEIN (FRAGM	3.05e-03
29	70	85.4	347 14	Q9WFJ1	POL POLYPROTEIN (FRAGM	3.05e-03
30	70	85.4	347 14	Q9WBX4	POL POLYPROTEIN (FRAGM	3.05e-03
31	70	85.4	347 14	Q9WBX5	POL POLYPROTEIN (FRAGM	3.05e-03
32	70	85.4	347 14	Q9YLO7	POL POLYPROTEIN (FRAGM	3.05e-03
33	70	85.4	347 14	Q9Z066	POL PROTEIN (FRAGMENT)	3.05e-03
34	70	85.4	347 14	Q9WF11	POL POLYPROTEIN (FRAGM	3.05e-03
35	70	85.4	347 14	Q9WBX3	POL POLYPROTEIN (FRAGM	3.05e-03
36	70	85.4	347 14	Q9WBX2	POL POLYPROTEIN (FRAGM	3.05e-03
37	70	85.4	347 14	Q9YLO4	POL POLYPROTEIN (FRAGM	3.05e-03
38	70	85.4	347 14	Q9YLO5	POL POLYPROTEIN (FRAGM	3.05e-03
39	70	85.4	367 14	Q9YJY5	POL POLYPROTEIN (FRAGM	3.05e-03
40	70	85.4	402 14	Q9YJY1	POL POLYPROTEIN (FRAGM	3.05e-03
41	70	85.4	403 14	Q9YJY4	POL POLYPROTEIN (FRAGM	3.05e-03
42	70	85.4	1000 14	Q9WF59	HIV-1 ISOLATE C-96BW01	3.05e-03
43	70	85.4	1001 14	Q9WSP0	POL PROTEIN (FRAGMENT)	3.05e-03
44	70	85.4	1003 14	Q9WSP9	POL PROTEIN (FRAGMENT)	3.05e-03
45	70	85.4	1427 14	Q9WF62	HIV-1 ISOLATE C-96BW01	3.05e-03

ALIGNMENTS

RESULT 1  
ID Q75826 PRELIMINARY; PRT; 245 AA.  
AC Q75832;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)  
DE REVERSE TRANSCRIPTASE (FRAGMENT).  
GN POL.  
OS Human immunodeficiency virus type 1.  
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-WP58 FROM AUSTRALIA;  
RA ZHENG N.N., HURREN L., NEILAN B.A., COOPER D.A., DELANEY S.F.,  
RA MCQUEEN P.W.;  
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.  
DR EMBL; U64183; AAB05326.1; -.  
DR HSSP; P04585; 1RT2.  
DR PFAM; PF00078; IRT.1.  
KW RNA-directed DNA polymerase.  
FT NON\_TER 1 1  
FT NON\_TER 245 245  
SQ SEQUENCE 245 AA; 28738 MW; 48907759 CRC32;

Query Match 100.08; Score 82; DB 14; Length 245;  
Best Local Similarity 100.08; Pred. No. 5.06e-06;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 9 YPKVKQWPL 17  
| | | | | | | | | |  
Qy 1 YPKVKQWPL 9

RESULT 2						
ID Q75832	PRELIMINARY;	PRT;	245 AA.			
AC Q75832;						
DT 01-NOV-1996	(TrEMBLrel. 01, Created)					
DT 01-NOV-1996	(TrEMBLrel. 01, Last sequence update)					
DT 01-NOV-1999	(TrEMBLrel. 12, Last annotation update)					
DE REVERSE TRANSCRIPTASE	(FRAGMENT).					
GN POL.						
OS Human immunodeficiency virus type 1.						
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.						
RN [1]						

THIS PAGE BLANK (USPTO)

QY 2 PKVKQWPL 9  
Search completed: Fri Apr 14 23:07:58 2000  
Job time : 42 secs.

DR PDB; 1HVP; 15-APR-92.  
DR PDB; 9HVP; 15-JUL-92.  
DR PDB; 1HRH; 15-OCT-94.  
DR PDB; 3PHV; 15-JAN-92.  
DR PDB; 2HMI; 14-OCT-98.  
DR PDB; 1HOS; 31-OCT-93.  
DR PDB; 3HVT; 15-OCT-94.  
DR PDB; 1HVI; 30-APR-94.  
DR PDB; 1HVJ; 30-APR-94.  
DR PDB; 1HVX; 30-APR-94.  
DR PDB; 1HVL; 30-APR-94.  
DR PDB; 1HEF; 31-MAY-94.  
DR PDB; 1HEG; 31-MAY-94.  
DR PDB; 1HMV; 31-MAR-95.  
DR PDB; 1HNI; 03-JUN-95.  
DR PDB; 1HNV; 10-JUL-95.  
DR PDB; 1HPS; 31-AUG-94.  
DR PDB; 1HTE; 31-JUL-94.  
DR PDB; 1HTF; 31-JUL-94.  
DR PDB; 1HTG; 31-JUL-94.  
DR PDB; 1SBG; 15-OCT-94.  
DR PDB; 1DLO; 01-AUG-96.  
DR PDB; 1GNN; 08-NOV-96.  
DR PDB; 1GNO; 08-NOV-96.  
DR PDB; 1AJV; 20-AUG-97.  
DR PDB; 1AJX; 17-SEP-97.  
DR PDB; 1RVL; 07-FEB-95.  
DR PDB; 1RVM; 07-FEB-95.  
DR PDB; 1RVN; 07-FEB-95.  
DR PDB; 1RVO; 07-FEB-95.  
DR PDB; 1RVP; 07-FEB-95.  
DR PDB; 1RVO; 15-MAY-95.  
DR PDB; 1RVR; 15-MAY-95.  
DR PDB; 1MER; 15-APR-98.  
DR PDB; 1MES; 15-APR-98.  
DR PDB; 1MET; 15-APR-98.  
DR PDB; 1MEU; 15-APR-98.  
DR PDB; 1BQM; 16-FEB-99.  
DR PDB; 1BQN; 16-FEB-99.  
DR HIV; M15654; POL\$BH102.  
DR PROSITE; PS00141; ASP\_PROTEASE; 1.  
DR PFAM; PF00075; rnaaseH; 1.  
DR PFAM; PF00077; rvp; 1.  
DR PFAM; PF00078; rvt; 1.  
DR PFAM; PF00552; integrase; 1.  
DR PFAM; PF00665; rve; 1.  
KW AIDS; Polyprotein; Hydrolase; Aspartyl protease; Endonuclease;  
Nuclease; Transferase; RNA-directed DNA polymerase; 3D-structure.  
FT CHAIN 69 167  
FT ACT\_SITE 93 93  
FT STRAND 70 71  
FT STRAND 78 83  
FT TURN 84 85  
FT STRAND 86 92  
FT TURN 94 95  
FT STRAND 100 101  
FT STRAND 111 117  
FT TURN 118 119  
FT STRAND 120 134  
FT TURN 135 136  
FT STRAND 137 145  
FT STRAND 152 153  
FT HELIX 155 158  
FT TURN 159 162  
FT STRAND 164 166  
SQ SEQUENCE 1015 AA; 26F6A003 CRC32;

Query Match 85.4%; Score 70; DB 1; Length 1015;  
Best Local Similarity 100.0%; Pred No. 1.88e-03;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Db 186 PKVKQWPL 193  
|||||||



DR EMBL; X01762; -, NOT\_ANNOTATED\_CDS.  
DR PIR; A03967; GNVWYL.  
DR HSSP; P04585; 1REV.  
DR HIV; K02083; POLSPV22.  
DR PROSITE; PS00141; ASP\_PROTEASE; 1.  
DR PFAM; PF00075; InaseH; 1.  
DR PFAM; PF00077; rvp; 1.  
DR PFAM; PF00078; rvt; 1.  
DR PFAM; PF00552; Integrase; 1.  
DR PFAM; PF00665; rve; 1.  
KW AIDS; Polyprotein; Hydrolase; Aspartyl protease; Endonuclease;  
KW Nuclease; Transferase; RNA-directed DNA polymerase.  
FT CHAIN 69 167  
FT ACT\_SITE 93 93 BY SIMILARITY.  
SQ SEQUENCE 1015 AA; 115090 MW; 28A1FFC8 CRC32;  
  
Query Match 85.4%; Score 70; DB 1; Length 1015;  
Best Local Similarity 100.0%; Pred. No. 1.88e-03;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Db 186 PKVKQWPL 193  
QY 2 PKVKQWPL 9  
|||||||  
  
RESULT 15  
ID POL\_HV1B1 STANDARD; PRT; 1015 AA.  
AC P03366;  
DT 21-NOV-1986 (Rel. 01, Created)  
DT 01-JUL-1988 (Rel. 09, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE POL POLYPROTEIN [CONTAINS: PROTEASE (RETROPEPSIN) (EC 3.4.23.16);  
DE REVERSE TRANSCRIPTASE (EC 2.7.7.49); RIBONUCLEASE H (EC 3.1.26.4)].  
GN POL.  
OS Human immunodeficiency virus type 1 (BH10 isolate) (HIV-1).  
OC Viruses; Retroviral viruses; Retroviridae; Lentivirus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 8511123.  
RA RATNER L., HASELTINE W., PATARCA R., LIVAK K.J., STARGICH B.R.,  
RA JOSEPHS S.F., DORAN E.R., RAFALSKI J.A., WHITEHORN E.A.,  
RA BAUMEISTER K., IVANOFF L., PETTEWAY S.R. JR., PEARSON M.L.,  
RA LAUTENBERGER J.A., PAPAS T.S., GHARAYEB J., CHANG N.T., GALLO R.C.,  
RA WONG-STAAAL F.;  
RT "Complete nucleotide sequence of the AIDS virus, HTLV-III.";  
RL Nature 313:277-284(1985).  
RN [2]  
RP 3D-STRUCTURE MODELING OF PROTEASE DOMAIN.  
RX MEDLINE; 89146134.  
RA WEBER I.T., MILLER M., JASKOLSKI M., LEIS J., SKALKA A.M.,  
RA WLODAWER A.;  
RT "Molecular modeling of the HIV-1 protease and its substrate binding  
RT site.";  
RL Science 243:928-931(1989).  
RN [3]  
RP X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS) OF 69-167.  
RX MEDLINE; 90044107.  
RA LAPATTO R., BLUNDELL T., HEMMINGS A., OVERINGTON J., WILDERSPIN A.,  
RA WOOD S., MERSON J.R., WHITTLE P.J., DANLEY D.E., GEOGHEGAN K.F.,  
RA HAWRYLIK S.J., LEE S.E., SCHELD K.G., HOBART P.M.;  
RT "X-ray analysis of HIV-1 proteinase at 2.7-A resolution confirms  
RT structural homology among retroviral enzymes.";  
RL Nature 342:299-302(1989).  
RN [4]  
RP X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS) OF 69-167.  
RX MEDLINE; 90341771.  
RA BRICKSON J., NEIDHART D.J., VANDRIE J., KEMPF D.J., WANG X.C.,  
RA NORBECK D.W., PLATNER J.J., RITTENHOUSE J.W., TURON M., WIDEBURG N.,  
RA KOHLBRENNER W.E., SIMMER R., HELFRICH R., PAUL D.A., KNIGGE M.;  
RT "Design, activity, and 2.8 A crystal structure of a C2 symmetric  
RT inhibitor complexed to HIV-1 protease.";  
RL Science 249:527-533(1990).  
RN [5]  
  
RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS) OF 594-729.  
RX MEDLINE; 91188281.  
RA DAVIES J.F. II, HOSTOMSKA Z., HOSTOMSKY Z., JORDAN S.R.,  
RA MATTHEWS D.A.;  
RT "Crystal structure of the ribonuclease H domain of HIV-1 reverse  
RT transcriptase.";  
RL Science 252:88-95(1991).  
RN [6]  
RP X-RAY CRYSTALLOGRAPHY (3.0 ANGSTROMS) OF 168-723.  
RX MEDLINE; 93317673.  
RA JACOBO-MOLINA A., DING J., NANNI R.G., CLARK A.D. JR., LU X.,  
RA TANTILLO C., WILLIAMS R.L., KAWER G., FERRIS A.L., CLARK P., HIZI A.,  
RA HUGHES S.H., ARNOLD E.;  
RT "Crystal structure of human immunodeficiency virus type 1 reverse  
RT transcriptase complexed with double-stranded DNA at 3.0-A resolution  
RT shows bent DNA.";  
RL Proc. Natl. Acad. Sci. U.S.A. 90:6320-6324(1993).  
RN [7]  
RP X-RAY CRYSTALLOGRAPHY (3.5 ANGSTROMS) OF 168-723.  
RX MEDLINE; 92311658.  
RA KOHLSTADT L.A., WANG J., FRIEDMAN J.M., RICE P.A., STEITZ T.A.;  
RT "Crystal structure at 3.5-A resolution of HIV-1 reverse transcriptase  
RT complexed with an inhibitor.";  
RL Science 256:1783-1790(1992).  
RN [8]  
RP X-RAY CRYSTALLOGRAPHY (3.2 ANGSTROMS) OF 168-727.  
RX MEDLINE; 95166801.  
RA ROGERS D.W., GAMBLIN S.J., HARRIS B.A., RAY S., CULP J.S.,  
RA HELMIG B., WOOLF D.J., DEBOUCK C., HARRISON S.C.;  
RT "The structure of unliganded reverse transcriptase from the human  
RT immunodeficiency virus type 1.";  
RL Proc. Natl. Acad. Sci. U.S.A. 92:1222-1226(1995).  
RN [9]  
RP X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS) OF 168-725.  
RX MEDLINE; 95338599.  
RA DING J., DAS K., TANTILLO C., ZHANG W., CLARK A.D. JR., JESSEN S.,  
RA LU X., HSIOU Y., JACOBO-MOLINA A., ANDRIES K., ET A.L.;  
RT "Structure of HIV-1 reverse transcriptase in a complex with the non-  
RT nucleoside inhibitor alpha-APA R 95845 at 2.8-A resolution.";  
RL Structure 3:365-379(1995).  
RN [10]  
RP X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS) OF 168-723.  
RX MEDLINE; 96434330.  
RA HSIOU Y., DING J., DAS K., CLARK A.D. JR., HUGHES S.H., ARNOLD E.;  
RT "Structure of unliganded HIV-1 reverse transcriptase at 2.7-A  
RT resolution: Implications of conformational changes for polymerization  
RT and inhibition mechanisms.";  
RL Structure 4:853-860(1996).  
RN [11]  
RP X-RAY CRYSTALLOGRAPHY (3.1 ANGSTROMS) OF 168-723.  
RX MEDLINE; 99033049.  
RA HSIOU Y., DAS K., DING J., CLARK A.D. JR., KLEIM J.P., ROSNER M.,  
RA WINKLER I., REISS G., HUGHES S.H., ARNOLD E.;  
RT "Structures of Tyr188Leu mutant and wild-type HIV-1 reverse  
RT transcriptase complexed with the non-nucleoside inhibitor HBY 097:  
RT inhibitor flexibility is a useful design feature for reducing drug  
RT resistance.";  
RL J. Mol. Biol. 284:313-323(1998).  
RN [12]  
CC -1- PTM: CLEAVAGE SITES THAT YIELD THE MATURE PROTEINS REMAIN TO BE  
CC DETERMINED.  
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY A2; ALSO  
CC KNOWN AS THE RETROPEPSIN FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; M15654; AAA44198.1; -  
DR PIR; A03985; GNVWHL3.

CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; M38429; AAB03745.1; -;  
DR HSP; P03366; LHMV.  
DR HIV; M38429; POL\$JRCSE.  
DR PROSITE; PS00141; ASP\_PROTEASE; 1.  
DR PFAM; PF00075; inaseH; 1.  
DR PFAM; PF00077; rvp; 1.  
DR PFAM; PF00078; rvt; 1.  
DR PFAM; PF00552; integrase; 1.  
DR PFAM; PF00665; rve; 1.  
KW AIDS; Polyprotein; Hydrolase; Aspartyl protease; Endonuclease;  
KW Nuclease; Transferase; RNA-directed DNA polymerase.  
FT CHAIN 61 159  
FT ACT\_SITE 85 85 BY SIMILARITY.  
SQ SEQUENCE 1007 AA; 114081 MW; 492C03ED CRC32;  
  
Query Match 85.4%; Score 70; DB 1; Length 1007;  
Best Local Similarity 100.0%; Pred. No. 1.88e-03;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Db 178 PKVKQWPL 185  
|||||  
QY 2 PKVKQWPL 9  
  
RESULT 13  
ID POL\_HV1BR STANDARD; PRT; 1015 AA.  
AC P03367;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 01-NOV-1988 (Rel. 09, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE POL POLYPROTEIN [CONTAINS: PROTEASE (RETROPEPSIN) (EC 3.4.23.16);  
DE REVERSE TRANSCRIPTASE (EC 2.7.7.49); RIBONUCLEASE H (EC 3.1.26.4)].  
GN POL.  
OS Human immunodeficiency virus type 1 (BRU isolate) (HIV-1).  
OC Viruses; Retrovirdae; Retroviridae; Lentivirus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 85099333.  
RA ALIZON M., WAIN-HOBSON S., SONIGO P., DANOS O., COLE S., ALIZON M.;  
RT "Nucleotide sequence of the AIDS virus, LAV.";  
RL Cell 40:9-17(1985).  
RN [2]  
RP REVISIONS TO 23-35.  
RX MEDLINE; 86245056.  
RA ALIZON M., WAIN-HOBSON S., MONTAGNIER L., SONIGO P.;  
RT "Genetic variability of the AIDS virus; nucleotide sequence analysis  
RT of two isolates from African patients.";  
RL Cell 46:63-74(1986).  
RN [3]  
RP X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS).  
RX MEDLINE; 92190341.  
RA SPINELLI S., LIU Q.Z., ALZARI P.M., HIREL P.H., POLJAK R.J.;  
RT "The three-dimensional structure of the aspartyl protease from the  
RT HIV-1 isolate BRU.";  
RL Biochimie 73:1391-1396(1991).  
CC -1- PTM: CLEAVAGE SITES THAT YIELD THE MATURE PROTEINS REMAIN TO BE  
CC DETERMINED.  
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY A2; ALSO  
CC KNOWN AS THE RETROPEPSIN FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; K02013; -; NOT\_ANNOTATED\_CDS.  
DR PIR; A03966; GNVWLV.  
DR PDB; 1HHP; 15-OCT-92.  
DR PROSITE; PS00141; ASP\_PROTEASE; 1.  
DR PFAM; PF00075; inaseH; 1.  
DR PFAM; PF00077; rvp; 1.  
DR PFAM; PF00078; rvt; 1.  
DR PFAM; PF00552; integrase; 1.  
DR PFAM; PF00665; rve; 1.  
KW AIDS; Polyprotein; Hydrolase; Aspartyl protease; Endonuclease;  
KW Nuclease; Transferase; RNA-directed DNA polymerase; 3D-structure.  
FT CHAIN 69 167  
FT ACT\_SITE 93 93 BY SIMILARITY.  
FT STRAND 78 82  
FT STRAND 87 92  
FT TURN 94 95  
FT STRAND 100 102  
FT STRAND 111 117  
FT TURN 118 119  
FT STRAND 120 134  
FT TURN 135 136  
FT STRAND 137 145  
FT STRAND 152 153  
FT HELIX 155 161  
FT TURN 162 162  
SQ SEQUENCE 1015 AA; 115031 MW; F34CS47E CRC32;  
  
Query Match 85.4%; Score 70; DB 1; Length 1015;  
Best Local Similarity 100.0%; Pred. No. 1.88e-03;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Db 186 PKVKQWPL 193  
|||||  
QY 2 PKVKQWPL 9  
  
RESULT 14  
ID POL\_HV1PV STANDARD; PRT; 1015 AA.  
AC P03368;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 01-NOV-1988 (Rel. 09, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE POL POLYPROTEIN [CONTAINS: PROTEASE (RETROPEPSIN) (EC 3.4.23.16);  
DE REVERSE TRANSCRIPTASE (EC 2.7.7.49); RIBONUCLEASE H (EC 3.1.26.4)].  
GN POL.  
OS Human immunodeficiency virus type 1 (PV22 isolate) (HIV-1).  
OC Viruses; Retrovirdae; Retroviridae; Lentivirus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 85111157.  
RA MUESING M.A., SMITH D.H., CABRADILLA C.D., BENTON C.V., LASKY L.A.,  
RA CAPON D.J.;  
RT "Nucleic acid structure and expression of the human  
RT AIDS/lymphadenopathy retrovirus.";  
RL Nature 313:450-458(1985).  
RN [2]  
RP REVISION.  
RA MUESING M.A.;  
RL Submitted (XXX-1987) to the HIV data bank.  
CC -1- PTM: CLEAVAGE SITES THAT YIELD THE MATURE PROTEINS REMAIN TO BE  
CC DETERMINED.  
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY A2; ALSO  
CC KNOWN AS THE RETROPEPSIN FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; K02083; AAB59867.1; -;

GN POL.  
Human immunodeficiency virus type 1 (MN isolate) (HIV-1).  
OC Viruses; Retroviral viruses; Retroviridae; Lentivirus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 88219542.  
RA GURGO C., GUO H.-G., FRANCHINI G., ALDOVINI A., COLLALTI E.,  
RA FARRELL K., WONG-STAAAL F., GALLO R.C., REITZ M.S. JR.;  
RT "Envelope sequences of two new United States HIV-1 isolates.";  
RL Virology 164:531-536(1988).  
CC -1- PTM: CLEAVAGE SITES THAT YIELD THE MATURE PROTEINS REMAIN TO BE  
CC DETERMINED.  
CC -1- MISCELLANEOUS: THE MN ISOLATE WAS TAKEN FROM A PEDIATRIC AIDS  
CC PATIENT IN 1984.  
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY A2; ALSO  
CC KNOWN AS THE RETROPEPSIN FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC at the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).  
CC -----  
CC EMBL; M17449; ; NOT\_ANNOTATED\_CDS.  
DR HSSP; P03366; IRRV.  
DR HIV; M1749; POLSMN.  
DR PROSITE; PS00141; ASP\_PROTEASE; 1.  
DR PFAM; PF000075; rnaseh; 1.  
DR PFAM; PF000077; rvp; 1.  
DR PFAM; PF000078; rvt; 1.  
DR PFAM; PF00352; integrase; 1.  
DR PFAM; PF00665; rve; 1.  
CC AIDS; Polyprotein; Hydrolase; Aspartyl protease; Endonuclease;  
KW Nuclease; Transferase; RNA-directed DNA polymerase.  
FT CHAIN 60 158  
FT ACT\_SITE 84 84 BY SIMILARITY.  
FT SITE 565 565 IN-FRAME TERMINATION CODON.  
FT SEQUENCE 1006 AA; 113860 MW; 70477EC0 CRC32;  
Query Match 85.48; Score 70; DB 1; Length 1006;  
Best Local Similarity 100.08; Pred. No. 1.88e-03;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Db 177 PKVQWPL 184  
QY 2 PKVQWPL 9  
|||||||  
RESULT 12  
ID POL\_HVLJR STANDARD; PRT; 1007 AA.  
AC P20875;  
DT 01-FEB-1991 (Rel. 17, Created)  
DT 01-FEB-1991 (Rel. 17, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE POL POLYPROTEIN [CONTAINS: PROTEASE (RETROPEPSIN) (EC 3.4.23.16);  
DE REVERSE TRANSCRIPTASE (EC 2.7.7.49); RIBONUCLEASE H (EC 3.1.26.4)].  
GN POL.  
OS Human immunodeficiency virus type 1 (JRCF isolate) (HIV-1).  
OC Viruses; Retroviral viruses; Retroviridae; Lentivirus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA KOYANAGI S., CHEN I.S.Y.;  
RL Submitted (DEC-1998) to the HIV data bank.  
CC -1- PTM: CLEAVAGE SITES THAT YIELD THE MATURE PROTEINS REMAIN TO BE  
CC DETERMINED.  
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY A2; ALSO  
CC KNOWN AS THE RETROPEPSIN FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC at the European Bioinformatics Institute. There are no restrictions on its

KW AIDS; Polyprotein; Hydrolase; Aspartyl protease; Endonuclease;  
KW Nuclease; Transferase; RNA-directed DNA polymerase; 3D-structure.  
FT CHAIN 57 155  
FT ACT\_SITE 81 81  
FT STRAND 66 71  
FT TURN 72 73  
FT STRAND 74 80  
FT STRAND 87 90  
FT STRAND 98 103  
FT TURN 106 107  
FT STRAND 110 115  
FT TURN 119 122  
FT STRAND 123 124  
FT STRAND 125 128  
FT STRAND 131 134  
FT STRAND 140 141  
FT HELIX 143 149  
FT TURN 150 150  
SQ SEQUENCE 1003 AA; 113723 MW; A94EB76C CRC32;  
Query Match 85.4%; Score 70; DB 1; Length 1003;  
Best Local Similarity 100.0%; Pred. No. 1.88e-03;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Db 174 PKVKQWPL 181  
QY 2 PKVKQWPL 9  
RESULT 9  
ID POL\_HVIY2 STANDARD; PRT; 1003 AA.  
AC P35963;  
DT 01-JUN-1994 (Rel. 29, Created)  
DT 01-JUN-1994 (Rel. 29, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE POL POLYPROTEIN [CONTAINS: PROTEASE (RETROPEPSIN) (EC 3.4.23.16);  
DE REVERSE TRANSCRIPTASE (EC 2.7.7.49); RIBONUCLEASE H (EC 3.1.26.4)].  
GN POL.  
OS Human immunodeficiency virus type 1 (YU-2 isolate) (HIV-1).  
OC Viruses; Retroviral viruses; Retroviridae; Lentivirus.  
RN [1]  
RX MEDLINE; 93021387.  
RA LI Y., HUI H., BURGESS C.J., PRICE R.W., SHARP P.M., HAHN B.H.,  
RA "Complete nucleotide sequence, genome organization, and biological  
RT properties of human immunodeficiency virus type 1 in vivo: evidence  
RT for limited defectiveness and complementation.";  
RL J. Virol. 66:6587-6600(1992).  
CC -1- PTM: CLEAVAGE SITES THAT YIELD THE MATURE PROTEINS REMAIN TO BE  
CC DETERMINED.  
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY A2; ALSO  
CC KNOWN AS THE RETROPEPSIN FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: M93258; -; NOT\_ANNOTATED\_CDS.  
DR PIR: B44001; B44001.  
DR HSP: P04585; 1RTH.  
DR PROSITE: PS00141; ASP\_PROTEASE; 1.  
DR PFAM: PF00075; rnsase; 1.  
DR PFAM: PF00077; rvp; 1.  
DR PFAM: PF00078; rvt; 1.  
DR PFAM: PF00552; integrase; 1.  
DR PFAM: PF00663; rve; 1.  
KW AIDS; Polyprotein; Hydrolase; Aspartyl protease; Endonuclease;  
KW Nuclease; Transferase; RNA-directed DNA polymerase.

FT CHAIN 57 155  
FT ACT\_SITE 82 82  
SQ SEQUENCE 1003 AA; 113794 MW; 99272DF9 CRC32;  
Query Match 85.4%; Score 70; DB 1; Length 1003;  
Best Local Similarity 100.0%; Pred. No. 1.88e-03;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Db 174 PKVKQWPL 181  
QY 2 PKVKQWPL 9  
RESULT 10  
ID POL\_HV1H2 STANDARD; PRT; 1003 AA.  
AC P04585; O09777;  
DT 13-AUG-1987 (Rel. 05, Created)  
DT 15-JUL-1999 (Rel. 38, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE POL POLYPROTEIN [CONTAINS: PROTEASE (RETROPEPSIN) (EC 3.4.23.16);  
DE REVERSE TRANSCRIPTASE (EC 2.7.7.49); RIBONUCLEASE H (EC 3.1.26.4)].  
GN POL.  
OS Human immunodeficiency virus type 1 (HXB2 isolate) (HIV-1).  
OC Viruses; Retroviral viruses; Retroviridae; Lentivirus.  
RN [1]  
RX MEDLINE; 87299196.  
RA RATNER L., FISHER A., JAGODZINSKI L.L., MITSUYA H., LIOU R.-S.,  
RA GALLO R.C., WONG-STAAL F.;  
RA "Complete nucleotide sequences of functional clones of the AIDS  
RT virus.";  
RL AIDS Res. Hum. Retroviruses 3:57-69(1987).  
RN [2]  
RP REVISIONS.  
RA OGATA N., ALTER H.J., MILLER R.H., PURCELL R.H.;  
RA Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 57-155.  
RX MEDLINE; 99043699.  
RA KERVINEN J., LUBKOWSKI J., ZDANOV A., BHATT D., DUNN B.M., HUI K.Y.,  
RA POWELL D.J., KAY J., WLODAR A., GUSTCHINA A.;  
RT "Toward a universal inhibitor of retroviral proteases: comparative  
RT analysis of the interactions of LP-130 complexed with proteases from  
RT HIV-1, FIV, and ERAV.";  
RL Protein Sci. 7:2314-2323(1998).  
RN [4]  
RP STRUCTURE BY NMR OF 57-155.  
RX MEDLINE; 97022126.  
RA YAMAZAKI T., HINCK A.P., WANG Y.-X., NICHOLSON L.K., TORCHIA D.A.,  
RA WINGFIELD P., STAHL S.J., KAUFMAN J.D., CHANG C.-H., DOMAILLE P.J.,  
RA LAM P.Y.S.;  
RT "Three-dimensional solution structure of the HIV-1 protease complexed  
RT with DMP323, a novel cyclic urea-type inhibitor, determined by  
RT nuclear magnetic resonance spectroscopy.";  
RL Protein Sci. 5:495-508(1996).  
RN [5]  
RP X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS) OF 156-595.  
RX MEDLINE; 96097398.  
RA REN J., ESNOUF R.M., HOPKINS A.L., ROSS C., JONES E.Y., STAMMERS D.K.,  
RA STUART D.I.;  
RT "The structure of HIV-1 reverse transcriptase complexed with  
RT 9-chloro-TIBO: lessons for inhibitor design.";  
RL Structure 3:915-926(1995).  
RN [6]  
RP X-RAY CRYSTALLOGRAPHY (2.55 ANGSTROMS) OF 156-595.  
RX MEDLINE; 96208551.  
RA HOPKINS A.L., REN J., ESNOUF R.M., WILLCOX B.E., JONES E.Y., ROSS C.,  
RA MIYASAKA T., WALKER R.T., TANAKA H., STAMMERS D.K., STUART D.I.;  
RT "Complexes of HIV-1 reverse transcriptase with inhibitors of the HEPT  
RT series reveal conformational changes relevant to the design of potent  
RT non-nucleoside inhibitors.";  
RL J. Med. Chem. 39:1589-1600(1996).  
RN [7]

FT STRAND 125 134  
 FT STRAND 140 141  
 FT HELIX 143 149  
 FT TURN 150 150  
 FT STRAND 152 154  
 SQ SEQUENCE 1003 AA; 113535 MW; 5ED59879 CRC32;

Query Match 85.4%; Score 70; DB 1; Length 1003;  
 Best Local Similarity 100.0%; Pred. No. 1.88e-03;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 174 PKVKQWPL 181  
 |||||  
 QY 2 PKVKQWPL 9

RESULT 7  
 ID POL\_HV10Y STANDARD; PRT; 1003 AA.  
 AC P20892;  
 DT 01-FEB-1991 (Rel. 17, Created)  
 DT 01-FEB-1991 (Rel. 17, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE POL POLYPROTEIN [CONTAINS: PROTEASE (RETROPEPSIN) (EC 3.4.23.16);  
 DE REVERSE TRANSCRIPTASE (EC 2.7.7.49); RIBONUCLEASE H (EC 3.1.26.4)].  
 GN POL.  
 OS Human immunodeficiency virus type 1 (OVI isolate) (HIV-1).  
 OC Viruses; Retroviral viruses; Retroviridae; Lentivirus.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE; 90148544.  
 RA HUET T., DAZZA M.C., BRUN-VEZINET F., ROELANTS G.E., WAIN-HOBSON S.;  
 RT "A highly defective HIV-1 strain isolated from a healthy Gabonese  
 RT individual presenting an atypical western blot.";  
 RL AIDS 3:707-715(1989).  
 CC -1- PTM: CLEAVAGE SITES THAT YIELD THE MATURE PROTEINS REMAIN TO BE  
 CC DETERMINED.  
 CC -1- MISCELLANEOUS: THE OVI ISOLATE WAS TAKEN FROM THE BLOOD OF A  
 CC HEALTHY GABONESE INDIVIDUAL.  
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY A2; ALSO  
 CC KNOWN AS THE RETROPEPSIN FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; M26727; AAA83392.1; -  
 CC HSP; P03366; LRVR.  
 CC HIV; M26727; POL\$OVI.  
 CC PROSITE; PS00141; ASP\_PROTEASE; 1.  
 CC PFAM; PF00075; rnaseH; 1.  
 CC PFAM; PF00077; rvp; 1.  
 CC PFAM; PF00078; rvt; 1.  
 CC PFAM; PF00552; integrase; 1.  
 CC PFAM; PF00665; rve; 1.  
 CC AIDS; Polyprotein; Hydrolase; Aspartyl protease; Endonuclease;  
 KW Nuclease; Transferase; RNA-directed DNA polymerase.  
 FT CHAIN 57 155  
 FT ACT\_SITE 81 81  
 FT ACT\_SITE BY SIMILARITY.  
 SQ SEQUENCE 1003 AA; 113718 MW; E50B705E CRC32;

Query Match 85.4%; Score 70; DB 1; Length 1003;  
 Best Local Similarity 100.0%; Pred. No. 1.88e-03;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 174 PKVKQWPL 181  
 |||||  
 QY 2 PKVKQWPL 9

RESULT 8  
 ID POL\_HV1A2 STANDARD; PRT; 1003 AA.  
 AC P03369;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 01-NOV-1988 (Rel. 09, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE POL POLYPROTEIN [CONTAINS: PROTEASE (RETROPEPSIN) (EC 3.4.23.16);  
 DE REVERSE TRANSCRIPTASE (EC 2.7.7.49); RIBONUCLEASE H (EC 3.1.26.4)].  
 GN POL.  
 OS Human immunodeficiency virus type 1 (ARV2/SF2 isolate) (HIV-1).  
 OC Viruses; Retroviral viruses; Retroviridae; Lentivirus.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE; 85090453.  
 RA SANCHEZ-PESCADOR R., POWER M.D., BARR P.J., STEIMER K.S.,  
 RA STEMPHEN M.M., BROWN-SHIMER S.L., GEE W.W., RENARD A., RANDOLPH A.,  
 RA LEVY J.A., DINA D., LUCIW P.A.;  
 RT "Nucleotide sequence and expression of an AIDS-associated retrovirus  
 RT (ARV-2)." ;  
 RL Science 227:484-492(1985).  
 RN [2]  
 RP X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS) OF 57-155.  
 RX MEDLINE; 89346747.  
 RA WLODAR A., MILLER M., JASKOLSKI M., SATHYANARAYANA B.K.,  
 RA BALDWIN E., WEBER I.T., SELK L.M., CLAWSON L., SCHNEIDER J.,  
 RA KENT S.B.H.;  
 RT "Conserved folding in retroviral proteases: crystal structure of a  
 RT synthetic HIV-1 protease." ;  
 RL Science 245:616-621(1989).  
 RN [3]  
 RP X-RAY CRYSTALLOGRAPHY (2.05 ANGSTROMS) OF 57-155 OF COMPLEX WITH INH.  
 RA ABBENANTE G., MARCH D.R., BERGMAN D.A., DANCER R., HUNT P.,  
 RA GARNHAM B., MARTIN J.L., FAIRLIE D.P.;  
 RL Submitted (OCT-1995) to the PDB data bank.  
 RN [4]  
 RP X-RAY CRYSTALLOGRAPHY (1.75 ANGSTROMS) OF 57-155 OF COMPLEX WITH INH.  
 RA MARCH D.R., ABBENANTE G., BERGMAN D.A., BRINKWORTH R.I.,  
 RA WICKRAMASINGHE W., BEGUN J., MARTIN J.L., FAIRLIE D.P.;  
 RL Submitted (FEB-1996) to the PDB data bank.  
 RN [5]  
 RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS) OF 57-155.  
 RX MEDLINE; 96438794.  
 RA ROSE R.B., CRAIK C.S., DOUGLAS N.L., STROUD R.M.;  
 RT "Three-dimensional structures of HIV-1 and SIV protease product  
 RT complexes." ;  
 RL Biochemistry 35:12933-12944(1996).  
 CC -1- PTM: CLEAVAGE SITES THAT YIELD THE MATURE PROTEINS REMAIN TO BE  
 CC DETERMINED.  
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY A2; ALSO  
 CC KNOWN AS THE RETROPEPSIN FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; K02007; AAB59876.1; -  
 CC PIR; A03968; GNVWAZ.  
 CC PDB; 3HVP; 15-JAN-90.  
 CC PDB; 1CPI; 08-MAR-96.  
 CC PDB; 1MTR; 01-AUG-96.  
 CC PDB; 1YTG; 12-MAR-97.  
 CC PDB; 1YTH; 12-MAR-97.  
 CC HIV; K02007; POL\$SF2.  
 CC PROSITE; PS00141; ASP\_PROTEASE; 1.  
 CC PFAM; PF00075; rnaseH; 1.  
 CC PFAM; PF00077; rvp; 1.  
 CC PFAM; PF00078; rvt; 1.  
 CC PFAM; PF00552; integrase; 1.  
 CC PFAM; PF00665; rve; 1.

CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; M27323; AAA44869.1; -  
DR PIR; JQ0067; GNLJND.  
DR HSSP; P03366; 1HMV.  
DR HIV; M27323; POL\$NDK.  
DR PROSITE; PS00141; ASP\_PROTEASE; 1.  
DR PFAM; PF00075; rnaseh; 1.  
DR PFAM; PF00077; rvp; 1.  
DR PFAM; PF00078; rvt; 1.  
DR PFAM; PF00552; integrase; 1.  
DR PFAM; PF00665; rve; 1.  
KW AIDS; Polyprotein; Hydrolase; Aspartyl protease; Endonuclease;  
KW Nuclease; Transferase; RNA-directed DNA polymerase.  
FT CHAIN 56 134  
FT ACT\_SITE 80 80 BY SIMILARITY.  
SQ SEQUENCE 1002 AA; 113621 MW; FC5DF15F CRC32;

Query Match 85.4%; Score 70; DB 1; Length 1002;  
Best Local Similarity 100.0%; Pred. No. 1.88e-03;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 173 PKVQWPL 180  
|||||||  
QY 2 PKVQWPL 9

RESULT 5  
ID POL\_HVIRH STANDARD; PRT: 1002 AA.  
AC P05959;  
DT 01-NOV-1988 (Rel. 09, Created)  
DT 01-NOV-1988 (Rel. 09, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE POL POLYPROTEIN [CONTAINS: PROTEASE (RETROPEPSIN) (EC 3.4.23.16);  
DE REVERSE TRANSCRIPTASE (EC 2.7.7.49); RIBONUCLEASE H (EC 3.1.26.4)].  
GN POL.  
OS Human immunodeficiency virus type 1 (RF/HAT isolate) (HIV-1).  
OC Viruses; Retroviridae; Lentivirus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA STARCICH B.R., HAHN B.H., SHAW G.M., MCNEELY P.D., MODROW S.,  
RA WOLF H., PARKS E.S., PARKS W.P., JOSEPHS S.F., GALLO R.C.,  
RA WONG-STAAAL F.  
RL Submitted (XXX-1987) to the HIV data bank.  
CC -!- PTM: CLEAVAGE SITES THAT YIELD THE MATURE PROTEINS REMAIN TO BE  
CC DETERMINED.

CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY A2; ALSO  
CC KNOWN AS THE RETROPEPSIN FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; M17451; AAA45053.1; -  
DR HSSP; P04585; 1RTH.  
DR HIV; M17451; POL\$RF.  
DR PROSITE; PS00141; ASP\_PROTEASE; 1.  
DR PFAM; PF00075; rnaseh; 1.  
DR PFAM; PF00077; rvp; 1.  
DR PFAM; PF00078; rvt; 1.  
DR PFAM; PF00552; integrase; 1.  
DR PFAM; PF00665; rve; 1.  
KW AIDS; Polyprotein; Hydrolase; Aspartyl protease; Endonuclease;  
KW Nuclease; Transferase; RNA-directed DNA polymerase.  
FT CHAIN 56 134  
FT ACT\_SITE 80 80 BY SIMILARITY.

SQ SEQUENCE 1002 AA; 113755 MW; 6DE2B1B2 CRC32;  
Query Match 85.4%; Score 70; DB 1; Length 1002;  
Best Local Similarity 100.0%; Pred. No. 1.88e-03;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 173 PKVQWPL 180  
|||||||  
QY 2 PKVQWPL 9

RESULT 6  
ID POL\_HVINS STANDARD; PRT: 1003 AA.  
AC P12497;  
DT 01-OCT-1989 (Rel. 12, Created)  
DT 01-OCT-1989 (Rel. 12, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE POL POLYPROTEIN [CONTAINS: PROTEASE (RETROPEPSIN) (EC 3.4.23.16);  
DE REVERSE TRANSCRIPTASE (EC 2.7.7.49); RIBONUCLEASE H (EC 3.1.26.4)].  
GN POL.  
OS Human immunodeficiency virus type 1 (New York-5 isolate) (HIV-1).  
OC Viruses; Retroviridae; Lentivirus.  
RN [1]  
RP SEQUENCE FROM N.A. (CLONE PNL4-3).  
RA BUCKLER C.E., BUCKLER-WHITE A.J., WILLEY R.L., MCCOY J.;  
RL Submitted (JUN-1988) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 57-155.  
RX MEDLINE; 90354401.  
RA FITZGERALD P.M.D., MCKEEVER B.M., VAN MIDDLESWORTH J.F.,  
RA SPRINGER J.P., HEIMBACH J.C., LEU C.-T., HERBER W.K., DIXON R.A.F.,  
RA DARKE P.L.;  
RT "Crystallographic analysis of a complex between human  
RT immunodeficiency virus type 1 protease and acetyl-pepstatin at 2.0-A  
RT resolution."  
RL J. Biol. Chem. 265:14209-14219(1990).  
CC -!- PTM: CLEAVAGE SITES THAT YIELD THE MATURE PROTEINS REMAIN TO BE  
CC DETERMINED.

CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY A2; ALSO  
CC KNOWN AS THE RETROPEPSIN FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; M19921; AAA44988.1; -  
DR PDB; 5HVP; 15-OCT-91.  
DR PDB; 4PHV; 31-OCT-93.  
DR HIV; M19921; POL\$NL43.  
DR PROSITE; PS00141; ASP\_PROTEASE; 1.  
DR PFAM; PF00075; rnaseh; 1.  
DR PFAM; PF00077; rvp; 1.  
DR PFAM; PF00078; rvt; 1.  
DR PFAM; PF00552; integrase; 1.  
DR PFAM; PF00665; rve; 1.  
KW AIDS; Polyprotein; Hydrolase; Aspartyl protease; Endonuclease;  
KW Nuclease; Transferase; RNA-directed DNA polymerase; 3D-structure.  
FT CHAIN 57 155  
FT ACT\_SITE 81 81  
FT STRAND 58 59  
FT STRAND 66 71  
FT TURN 72 73  
FT STRAND 74 80  
FT TURN 82 83  
FT STRAND 88 90  
FT STRAND 99 105  
FT TURN 106 107  
FT STRAND 108 122  
FT TURN 123 124

Query Match 85.4%; Score 70; DB 1; Length 1002;  
 Best Local Similarity 100.0%; Pred. No. 1.88e-03;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 173 PKVKQWPL 180  
 QY 2 PKVKQWPL 9  
 |||||

RESULT 2  
 ID POL\_HV122 STANDARD; PRT; 1002 AA.  
 AC P12499;  
 DT 01-OCT-1989 (Rel. 12, Last sequence update)  
 DT 01-OCT-1989 (Rel. 12, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE POL POLYPROTEIN [CONTAINS: PROTEASE (RETROPEPSIN) (EC 3.4.23.16);  
 DE REVERSE TRANSCRIPTASE (EC 2.7.7.49); RIBONUCLEASE H (EC 3.1.26.4)].  
 GN POL.  
 OS Human immunodeficiency virus type 1 (Z2/CDC-234 isolate) (HIV-1).  
 CC Viruses; Retrovirdae; Retroviridae; Lentivirus.  
 CC [1]  
 CC SEQUENCE FROM N.A.  
 RA THEODORE T., BUCKLER-WHITE A.;  
 RL Submitted (NOV-1989) to the HIV data bank.  
 CC -1- PTM: CLEAVAGE SITES THAT YIELD THE MATURE PROTEINS REMAIN TO BE  
 CC DETERMINED.  
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY A2; ALSO  
 CC KNOWN AS THE RETROPEPSIN FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC [1]

EMBL; M22639; AAA5366.1; -  
 DR HSP; P03366; IRVR.  
 DR HIV; M22639; POL\$2226.  
 DR PROSITE; PS00141; ASP\_PROTEASE; 1.  
 DR PFAM; PF00075; rnaaseh; 1.  
 DR PFAM; PF00077; rvp; 1.  
 DR PFAM; PF00078; rvt; 1.  
 DR PFAM; PF00552; integrase; 1.  
 DR PFAM; PF00665; rve; 1.  
 DR AIDS; Polyprotein; Hydrolase; Aspartyl protease; Endonuclease;  
 KW Nuclease; Transferase; RNA-directed DNA polymerase.  
 FT CHAIN 56 154  
 FT ACT\_SITE 81 81 BY SIMILARITY.  
 SQ SEQUENCE 1002 AA; 113724 MW; 8DAA803F CRC32;

Query Match 85.4%; Score 70; DB 1; Length 1002;  
 Best Local Similarity 100.0%; Pred. No. 1.88e-03;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 173 PKVKQWPL 180  
 QY 2 PKVKQWPL 9  
 |||||

RESULT 3  
 ID POL\_HV104 STANDARD; PRT; 1002 AA.  
 AC P24740;  
 DT 01-MAR-1992 (Rel. 21, Created)  
 DT 01-MAR-1992 (Rel. 21, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE POL POLYPROTEIN [CONTAINS: PROTEASE (RETROPEPSIN) (EC 3.4.23.16);  
 DE REVERSE TRANSCRIPTASE (EC 2.7.7.49); RIBONUCLEASE H (EC 3.1.26.4)].  
 GN POL.  
 OS Human immunodeficiency virus type 1 (strain Ugandan / isolate U455)  
 OS (HIV-1).  
 CC Viruses; Retrovirdae; Retroviridae; Lentivirus.

RN SEQUENCE FROM N.A.  
 RP MEDLINE; 91090981.  
 RX ORAM J.D., DOWNING R.G., ROFF M., CLEGG J.C.S., SERWADDA D.,  
 RA CARSWELL J.W.;  
 RA "Nucleotide sequence of a Ugandan HIV-1 provirus reveals genetic  
 RT diversity from other HIV-1 isolates.";  
 RL AIDS Res. Hum. Retroviruses 6:1073-1078(1990).  
 CC -1- PTM: CLEAVAGE SITES THAT YIELD THE MATURE PROTEINS REMAIN TO BE  
 CC DETERMINED.  
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY A2; ALSO  
 CC KNOWN AS THE RETROPEPSIN FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC [1]

EMBL; M62320; AAA75019.1; -  
 DR HSP; P03366; IRVR.  
 DR PROSITE; PS00141; ASP\_PROTEASE; 1.  
 DR PFAM; PF00075; rnaaseh; 1.  
 DR PFAM; PF00077; rvp; 1.  
 DR PFAM; PF00078; rvt; 1.  
 DR PFAM; PF00552; integrase; 1.  
 DR PFAM; PF00665; rve; 1.  
 DR AIDS; Polyprotein; Hydrolase; Aspartyl protease; Endonuclease;  
 KW Nuclease; Transferase; RNA-directed DNA polymerase.  
 FT CHAIN 56 154  
 FT ACT\_SITE 80 80 BY SIMILARITY.  
 SQ SEQUENCE 1002 AA; 113595 MW; D560B6FB CRC32;

Query Match 85.4%; Score 70; DB 1; Length 1002;  
 Best Local Similarity 100.0%; Pred. No. 1.88e-03;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 173 PKVKQWPL 180  
 QY 2 PKVKQWPL 9  
 |||||

RESULT 4  
 ID POL\_HV1ND STANDARD; PRT; 1002 AA.  
 AC P18802;  
 DT 01-NOV-1990 (Rel. 16, Created)  
 DT 01-NOV-1990 (Rel. 16, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE POL POLYPROTEIN [CONTAINS: PROTEASE (RETROPEPSIN) (EC 3.4.23.16);  
 DE REVERSE TRANSCRIPTASE (EC 2.7.7.49); RIBONUCLEASE H (EC 3.1.26.4)].  
 GN POL.  
 OS Human immunodeficiency virus type 1 (NDK isolate) (HIV-1).  
 CC Viruses; Retrovirdae; Retroviridae; Lentivirus.  
 CC [1]

SEQUENCE FROM N.A.  
 RX MEDLINE; 90034200.  
 RA SPIRE B., SIRE J., ZACHAR V., REY F., BARRE-SINOUSI F., GALIBERT F.,  
 RA HAMPE A., CHERMANN J.C.;  
 RA "Nucleotide sequence of HIV-1-NDK: a highly cytopathic strain of the  
 RT human immunodeficiency virus.";  
 RL Gene 81:275-284(1989).  
 CC -1- PTM: CLEAVAGE SITES THAT YIELD THE MATURE PROTEINS REMAIN TO BE  
 CC DETERMINED.  
 CC -1- MISCELLANEOUS: NDK, ISOLATED FROM A ZAIRIAN PATIENT AFFECTED WITH  
 CC AIDS, AND IS A HIGHLY CYTOPATHOGENIC STRAIN.  
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY A2; ALSO  
 CC KNOWN AS THE RETROPEPSIN FAMILY.  
 CC -----

This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its



\*\*\*\*\*

(TM)

\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Fri Apr 14 23:07:16 2000; MasPar time 6.20 Seconds  
Tabular output not generated. 43.365 Million cell updates/sec.

Title: >US-08-452-843-2  
Description: (1-9) from US08452843.pep  
Perfect Score: 82  
Sequence: 1 YPKVKQWPL 9

Scoring table: PAM 150  
Gap 15

Searched: 82229 seqs, 29864866 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: swiss-prot38  
1:swissprot

Statistics: Mean 25.222; Variance 32.943; scale 0.766

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	70	85.4	1002	1	POL_HV1EL	1.88e-03
2	70	85.4	1002	1	POL_HV1Z2	1.88e-03
3	70	85.4	1002	1	POL_HV1U4	1.88e-03
4	70	85.4	1002	1	POL_HV1ND	1.88e-03
5	70	85.4	1002	1	POL_HV1RH	1.88e-03
6	70	85.4	1002	1	POL_HV1N5	1.88e-03
7	70	85.4	1003	1	POL_HV1OY	1.88e-03
8	70	85.4	1003	1	POL_HV1A2	1.88e-03
9	70	85.4	1003	1	POL_HV1Y2	1.88e-03
10	70	85.4	1003	1	POL_HV1H2	1.88e-03
11	70	85.4	1006	1	POL_HV1M3	1.88e-03
12	70	85.4	1007	1	POL_HV1JR	1.88e-03
13	70	85.4	1015	1	POL_HV1BR	1.88e-03
14	70	85.4	1015	1	POL_HV1PV	1.88e-03
15	70	85.4	1015	1	POL_HV1B1	1.88e-03
16	70	85.4	1015	1	POL_HV1B5	1.88e-03
17	70	85.4	1027	1	POL_SIVC2	1.88e-03
18	67	81.7	1002	1	POL_HV1MA	8.59e-03
19	67	81.7	1035	1	POL_HV2KR	8.59e-03
20	65	79.3	1009	1	POL_SIVGB	2.32e-02
21	65	79.3	1038	1	POL_HV2D2	3.78e-02
22	64	78.0	147	1	YTI0_MVCTU	3.78e-02
23	64	78.0	1054	1	POL_SIVMK	3.78e-02

24	64	78.0	1056	1	POL_SIVM1	POL POLYPROTEIN [CONTA	3.78e-02
25	64	78.0	1057	1	POL_SIVAI	POL POLYPROTEIN [CONTA	3.78e-02
26	63	76.8	1124	1	POL_FIVT2	POL POLYPROTEIN [CONTA	6.15e-02
27	61	74.4	305	1	RNH_BPT4	RIBONUCLEASE H (EC 3.1	1.60e-01
28	61	74.4	1019	1	POL_SIVS4	POL POLYPROTEIN [CONTA	1.60e-01
29	61	74.4	1022	1	POL_SIVP5	POL POLYPROTEIN [CONTA	1.60e-01
30	61	74.4	1035	1	POL_HV2N2	POL POLYPROTEIN [CONTA	1.60e-01
31	61	74.4	1036	1	POL_HV2RO	POL POLYPROTEIN [CONTA	1.60e-01
32	61	74.4	1047	1	POL_SIVAI	POL POLYPROTEIN [CONTA	1.60e-01
33	61	74.4	1055	1	POL_HV2ST	POL POLYPROTEIN [CONTA	1.60e-01
34	61	74.4	1073	1	POL_HV2D1	POL POLYPROTEIN [CONTA	1.60e-01
35	61	74.4	1124	1	POL_FIVPE	POL POLYPROTEIN [CONTA	1.60e-01
36	61	74.4	1124	1	POL_FIVSD	POL POLYPROTEIN [CONTA	1.60e-01
37	61	74.4	1142	1	POL_HV2BE	POL POLYPROTEIN [CONTA	1.60e-01
38	60	73.2	1056	1	POL_BIV27	POL POLYPROTEIN [CONTA	2.57e-01
39	60	73.2	1056	1	POL_BIV06	POL POLYPROTEIN [CONTA	2.57e-01
40	58	70.7	229	1	GTH1_WHEAT	GLUTATHIONE S-TRANSFER	6.53e-01
41	58	70.7	291	1	GTH2_WHEAT	GLUTATHIONE S-TRANSFER	6.53e-01
42	58	70.7	1049	1	POL_HV2G1	POL POLYPROTEIN [CONTA	6.53e-01
43	58	70.7	1145	1	POL_EIAYV	POL POLYPROTEIN [CONTA	6.53e-01
44	58	70.7	1146	1	POL_EIAYV	POL POLYPROTEIN [CONTA	6.53e-01
45	58	70.7	1146	1	POL_EIAYV	POL POLYPROTEIN [CONTA	6.53e-01

## ALIGNMENTS

RESULT 1  
ID POL\_HV1EL STANDARD; PRT: 1002 AA.  
AC P04589; Q77906;  
DT 13-AUG-1987 (Rel. 05, Created)  
DT 01-NOV-1988 (Rel. 09, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE POL POLYPROTEIN [CONTAINS: PROTEASE (RETROPEPSIN) (EC 3.4.23.16);  
DE REVERSE TRANSCRIPTASE (EC 2.7.7.49); RIBONUCLEASE H (EC 3.1.26.4)].  
GN POL.  
OS Human immunodeficiency virus type 1 (ELI isolate) (HIV-1).  
OC Viruses; Retroviral viruses; Retroviridae; Lentivirus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 86245056.  
RA ALZON M., WAIN-HOBSON S., MONTAGNIER L., SONTIGO P.;  
RT "Genetic variability of the AIDS virus: nucleotide sequence analysis  
of two isolates from African patients."  
RL Cell 46:63-74(1986).  
CC -!- PTM: CLEAVAGE SITES THAT YIELD THE MATURE PROTEINS REMAIN TO BE  
CC DETERMINED.  
CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY A2; ALSO  
CC KNOWN AS THE RETROPEPSIN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: A07108; GAA00612.1; -;  
CC EMBL: K03454; AAA44325.1; ALT\_INT.  
CC HSSP: P03366; IHMV.  
CC HIV: K03454; POLSELI.  
CC PROSITE: PS00141; ASP\_PROTEASE; 1.  
CC PFAM: PF00075; rnaaseh; 1.  
CC PFAM: PF00077; rvp; 1.  
CC PFAM: PF00078; rvt; 1.  
CC PFAM: PF00552; Integrase; 1.  
CC PFAM: PF00665; rve; 1.  
CC AIDS; Polyprotein; Hydrolase; Aspartyl protease; Endonuclease;  
CC Nuclease; Transferase; RNA-directed DNA polymerase.  
FT CHAIN 56 154 PROTEASE.  
FT ACT\_SITE 80 80 BY SIMILARITY.  
SQ SEQUENCE 1002 AA; 114002 MW; 5DFEC55 CRC32;



THIS PAGE BLANK (USPTO)

TITLE RNA-directed DNA polymerase (EC 2.7.7.49) - human  
immunodeficiency virus type 1 (fragment)  
ALTERNATE\_NAMES reverse transcriptase  
ORGANISM #formal\_name human immunodeficiency virus type 1, HIV-1  
DATE 22-Nov-1993 #sequence\_revision 10-Nov-1995 #text\_change  
08-Sep-1997  
ACCESSIONS S32094  
REFERENCE S32047  
#authors Wildemann, B.; Haas, J.; Ehrhart, K.; Hahn, M.; Storch-Hagen,  
B.  
#submission submitted to the EMBL Data Library, January 1993  
#description In vivo comparison of zidovudine resistance in blood and CSF  
of HIV-1 infected patients.  
#accession S32094  
#molecule\_type DNA  
#residues 1-219 #label WIL  
#cross-references EMBL:X70626; NID:G287717; PID:G938223  
CLASSIFICATION #superfamily pol polyprotein  
KEYWORDS nucleotidyltransferase  
SUMMARY #length 219 #checksum 9992

Query Match 85.4%; Score 70; DB 2; Length 219;  
Best Local Similarity 100.0%; Pred. No. 1.27e-02;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 19 PKVKQWPL 26  
|||||||  
QY 2 PKVKQWPL 9

RESULT 14  
ENTRY S32157 #type fragment  
TITLE RNA-directed DNA polymerase (EC 2.7.7.49) - human  
immunodeficiency virus type 1 (fragment)  
ORGANISM #formal\_name human immunodeficiency virus type 1, HIV-1  
DATE 22-Nov-1993 #sequence\_revision 10-Nov-1995 #text\_change  
08-Sep-1997  
ACCESSIONS S32157  
REFERENCE S32117  
#authors Wildemann, B.; Haas, J.; Hahn, M.; Ehrhart, K.;  
Storch-Hagenlocher, B.  
#submission submitted to the EMBL Data Library, February 1993  
#description In vivo occurrence of drug resistance mutations under  
prolonged zidovudine treatment in HIV-1 infected patients  
with asymptomatic and advanced disease.

#accession S32157  
#molecule\_type DNA  
#residues 1-219 #label WIL  
#cross-references EMBL:X72318; NID:G288139; PID:G288140  
CLASSIFICATION #superfamily pol polyprotein  
KEYWORDS nucleotidyltransferase  
SUMMARY #length 219 #checksum 9685

Query Match 85.4%; Score 70; DB 2; Length 219;  
Best Local Similarity 100.0%; Pred. No. 1.27e-02;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 19 PKVKQWPL 26  
|||||||  
QY 2 PKVKQWPL 9

RESULT 15  
ENTRY S32078 #type fragment  
TITLE RNA-directed DNA polymerase (EC 2.7.7.49) - human  
immunodeficiency virus type 1 (fragment)  
ALTERNATE\_NAMES reverse transcriptase  
ORGANISM #formal\_name human immunodeficiency virus type 1, HIV-1  
DATE 22-Nov-1993 #sequence\_revision 10-Nov-1995 #text\_change  
08-Sep-1997

ACCESSIONS S32078  
REFERENCE S32047  
#authors Wildemann, B.; Haas, J.; Ehrhart, K.; Hahn, M.; Storch-Hagen,

#submission submitted to the EMBL Data Library, January 1993  
#description In vivo comparison of zidovudine resistance in blood and CSF  
of HIV-1 infected patients.  
#accession S32078  
#molecule\_type DNA  
#residues 1-219 #label WIL  
#cross-references EMBL:X70602; NID:G287701; PID:G938208  
CLASSIFICATION #superfamily pol polyprotein  
KEYWORDS nucleotidyltransferase  
SUMMARY #length 219 #checksum 9604

Query Match 85.4%; Score 70; DB 2; Length 219;  
Best Local Similarity 100.0%; Pred. No. 1.27e-02;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 19 PKVKQWPL 26  
|||||||  
QY 2 PKVKQWPL 9

Search completed: Fri Apr 14 23:06:58 2000  
Job time : 13 secs.

```
##molecule_type DNA
##residues 1-219 ##label WIL
##cross-references EMBL:X70610; NID:g287702; PID:g938209
CLASSIFICATION #superfamily pol polyprotein
KEYWORDS nucleotidyltransferase
SUMMARY #length 219 #checksum 281

Query Match 85.4%; Score 70; DB 2; Length 219;
Best Local Similarity 100.0%; Pred. No. 1.27e-02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 19 PKVKOWPL 26
QY 2 PKVKOWPL 9

RESULT 9
ENTRY S32152 #type fragment
TITLE RNA-directed DNA polymerase (EC 2.7.7.49) - human
ORGANISM immunodeficiency virus type 1 (fragment)
DATE #formal_name human immunodeficiency virus type 1, HIV-1
22-Nov-1993 #sequence_revision 10-Nov-1995 #text_change
08-Sep-1997
ACCESSIONS S32152
REFERENCE S32117
#authors Wildemann, B.; Haas, J.; Hahn, M.; Ehrhart, K.;
#submission Storch-Hagenlocher, B.
#description submitted to the EMBL Data Library, February 1993
#description In vivo occurrence of drug resistance mutations under
prolonged zidovudine treatment in HIV-1 infected patients
with asymptomatic and advanced disease.
#accession S32152
##molecule_type DNA
##residues 1-219 ##label WIL
##cross-references EMBL:X71097; NID:g288093; PID:g288094
CLASSIFICATION #superfamily pol polyprotein
KEYWORDS nucleotidyltransferase
SUMMARY #length 219 #checksum 105

Query Match 85.4%; Score 70; DB 2; Length 219;
Best Local Similarity 100.0%; Pred. No. 1.27e-02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 19 PKVKOWPL 26
QY 2 PKVKOWPL 9

RESULT 10
ENTRY S32159 #type fragment
TITLE RNA-directed DNA polymerase (EC 2.7.7.49) - human
ORGANISM immunodeficiency virus type 1 (fragment)
DATE #formal_name human immunodeficiency virus type 1, HIV-1
22-Nov-1993 #sequence_revision 10-Nov-1995 #text_change
08-Sep-1997
ACCESSIONS S32159
REFERENCE S32117
#authors Wildemann, B.; Haas, J.; Hahn, M.; Ehrhart, K.;
#submission Storch-Hagenlocher, B.
#description submitted to the EMBL Data Library, February 1993
#description In vivo occurrence of drug resistance mutations under
prolonged zidovudine treatment in HIV-1 infected patients
with asymptomatic and advanced disease.
#accession S32159
##molecule_type DNA
##residues 1-219 ##label WIL
##cross-references EMBL:X72317; NID:g288136; PID:g288137
CLASSIFICATION #superfamily pol polyprotein
KEYWORDS nucleotidyltransferase
SUMMARY #length 219 #checksum 9532

Query Match 85.4%; Score 70; DB 2; Length 219;
Best Local Similarity 100.0%; Pred. No. 1.27e-02;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 19 PKVKOWPL 26
QY 2 PKVKOWPL 9

RESULT 11
ENTRY S32160 #type fragment
TITLE RNA-directed DNA polymerase (EC 2.7.7.49) - human
ORGANISM immunodeficiency virus type 1 (fragment)
DATE #formal_name human immunodeficiency virus type 1, HIV-1
22-Nov-1993 #sequence_revision 10-Nov-1995 #text_change
08-Sep-1997
ACCESSIONS S32160
REFERENCE S32117
#authors Wildemann, B.; Haas, J.; Hahn, M.; Ehrhart, K.;
#submission Storch-Hagenlocher, B.
#description submitted to the EMBL Data Library, February 1993
#description In vivo occurrence of drug resistance mutations under
prolonged zidovudine treatment in HIV-1 infected patients
with asymptomatic and advanced disease.
#accession S32160
##molecule_type DNA
##residues 1-219 ##label WIL
##cross-references EMBL:X72319; NID:g288141; PID:g288142
CLASSIFICATION #superfamily pol polyprotein
KEYWORDS nucleotidyltransferase
SUMMARY #length 219 #checksum 9594

Query Match 85.4%; Score 70; DB 2; Length 219;
Best Local Similarity 100.0%; Pred. No. 1.27e-02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 19 PKVKOWPL 26
QY 2 PKVKOWPL 9

RESULT 12
ENTRY S32092 #type fragment
TITLE RNA-directed DNA polymerase (EC 2.7.7.49) - human
ALTERNATE_NAMES immunodeficiency virus type 1 (fragment)
ORGANISM reverse transcriptase
DATE #formal_name human immunodeficiency virus type 1, HIV-1
22-Nov-1993 #sequence_revision 10-Nov-1995 #text_change
08-Sep-1997
ACCESSIONS S32092
REFERENCE S32047
#authors Wildemann, B.; Haas, J.; Ehrhart, K.; Hahn, M.; Storch-Hagen,
B.
#submission submitted to the EMBL Data Library, January 1993
#description In vivo comparison of zidovudine resistance in blood and CSF
of HIV-1 infected patients.
#accession S32092
##molecule_type DNA
##residues 1-219 ##label WIL
##cross-references EMBL:X70624; NID:g287715; PID:g938221
CLASSIFICATION #superfamily pol polyprotein
KEYWORDS nucleotidyltransferase
SUMMARY #length 219 #checksum 1450

Query Match 85.4%; Score 70; DB 2; Length 219;
Best Local Similarity 100.0%; Pred. No. 1.27e-02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 19 PKVKOWPL 26
QY 2 PKVKOWPL 9

RESULT 13
ENTRY S32094 #type fragment
```

```
5
RESULT      5
ENTRY      S63753      #type fragment
TITLE      pol polyprotein - human immunodeficiency virus type 1
            (isolate RJ9533M) (fragment)
CONTAINS   retropepsin (EC 3.4.23.16); RNA-directed DNA polymerase (EC
            2.7.7.49)
ORGANISM   #formal_name human immunodeficiency virus type 1, HIV-1
            isolate RJ9533M
#variety   20-Jul-1996 #sequence_revision 27-Feb-1997 #text_change
DATE       08-Sep-1997
ACCESSIONS S63753; S63727
REFERENCE   S63731
#authors   Yamaguchi, K.
#submission submitted to the EMBL Data Library, July 1995
#accession S63753
#molecule_type DNA
#residues  1-160 #label YAM
##cross-references EMBL:U31409; NID:g961596; PID:g961597
##experimental_source isolate RJ9533M
REFERENCE   S63703
#authors   Yamaguchi, K.; Byrn, R.A.
#journal   Biochim. Biophys. Acta (1995) 1253:136-140
#title     Clinical isolates of HIV-1 contain few pre-existing
            proteinase inhibitor resistance-conferring mutations.
#cross-references MUID:96106422
#accession S63727
#status    nucleic acid sequence not shown
#molecule_type DNA
#residues  32-130 #label YAM
##cross-references EMBL:U31409
##experimental_source isolate RJ9533M
GENETICS
#gene      pol
CLASSIFICATION
#superfamily pol polyprotein
KEYWORDS   AIDS; aspartic proteinase; hydrolase; immunodeficiency;
            nucleotidyltransferase; polypeptide
FEATURE    32-130      #product retropepsin #status predicted #label RTP
SUMMARY    #length 160 #checksum 5613
            85.4%; Score 70; DB 2; Length 160;
            Best Local Similarity 100.0%; Pred. No. 1.27e-02;
            Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Query Match      85.4%; Score 70; DB 2; Length 160;
Best Local Similarity 100.0%; Pred. No. 1.27e-02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 149 PKVKQWPL 156
QY 2 PKVKQWPL 9

RESULT      6
ENTRY      S63732      #type fragment
TITLE      pol polyprotein - human immunodeficiency virus type 1
            (isolate RJ9435) (fragment)
CONTAINS   retropepsin (EC 3.4.23.16); RNA-directed DNA polymerase (EC
            2.7.7.49)
ORGANISM   #formal_name human immunodeficiency virus type 1, HIV-1
            isolate RJ9435
#variety   20-Jul-1996 #sequence_revision 27-Feb-1997 #text_change
DATE       08-Sep-1997
ACCESSIONS S63732; S63704
REFERENCE   S63731
#authors   Yamaguchi, K.
#submission submitted to the EMBL Data Library, July 1995
#accession S63732
#molecule_type DNA
#residues  1-162 #label YAM
##cross-references EMBL:U31386; NID:g961550; PID:g961551
##experimental_source isolate RJ9435
REFERENCE   S63703
#authors   Yamaguchi, K.; Byrn, R.A.
#journal   Biochim. Biophys. Acta (1995) 1253:136-140
#title     Clinical isolates of HIV-1 contain few pre-existing
```

```
proteinase inhibitor resistance-conferring mutations.
#cross-references MUID:96106422
#accession S63704
#status    nucleic acid sequence not shown
#molecule_type DNA
#residues  34-132 #label YAW
##cross-references EMBL:U31386
##experimental_source isolate RJ9435
GENETICS
#gene      pol
CLASSIFICATION
#superfamily pol polyprotein
KEYWORDS   AIDS; aspartic proteinase; hydrolase; immunodeficiency;
            nucleotidyltransferase; polypeptide
FEATURE    34-132      #product retropepsin #status predicted #label RTP
SUMMARY    #length 162 #checksum 3671
            85.4%; Score 70; DB 2; Length 162;
            Best Local Similarity 100.0%; Pred. No. 1.27e-02;
            Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Query Match      85.4%; Score 70; DB 2; Length 162;
Best Local Similarity 100.0%; Pred. No. 1.27e-02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 151 PKVKQWPL 158
QY 2 PKVKQWPL 9

RESULT      7
ENTRY      S32128      #type fragment
TITLE      RNA-directed DNA polymerase (EC 2.7.7.49) - human
            immunodeficiency virus type 1 (fragment)
ORGANISM   #formal_name human immunodeficiency virus type 1, HIV-1
            22-Nov-1993 #sequence_revision 10-Nov-1995 #text_change
DATE       08-Sep-1997
ACCESSIONS S32128
REFERENCE   S32117
#authors   Wildemann, B.; Haas, J.; Hahn, M.; Ehrhart, K.;
            Storch-Hagenlocher, B.
#submission submitted to the EMBL Data Library, February 1993
#description In vivo occurrence of drug resistance mutations under
            prolonged zidovudine treatment in HIV-1 infected patients
            with asymptomatic and advanced disease.
#accession S32128
#molecule_type DNA
#residues  1-219 #label WIL
##cross-references EMBL:X71094; NID:g287964; PID:g287965
CLASSIFICATION
#superfamily pol polyprotein
KEYWORDS   nucleotidyltransferase
SUMMARY    #length 219 #checksum 9883
            85.4%; Score 70; DB 2; Length 219;
            Best Local Similarity 100.0%; Pred. No. 1.27e-02;
            Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Query Match      85.4%; Score 70; DB 2; Length 219;
Best Local Similarity 100.0%; Pred. No. 1.27e-02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 19 PKVKQWPL 26
QY 2 PKVKQWPL 9

RESULT      8
ENTRY      S32079      #type fragment
TITLE      RNA-directed DNA polymerase (EC 2.7.7.49) - human
            immunodeficiency virus type 1 (fragment)
ALTERNATE_NAMES reverse transcriptase
ORGANISM   #formal_name human immunodeficiency virus type 1, HIV-1
            22-Nov-1993 #sequence_revision 10-Nov-1995 #text_change
DATE       08-Sep-1997
ACCESSIONS S32079
REFERENCE   S32047
#authors   Wildemann, B.; Haas, J.; Ehrhart, K.; Hahn, M.; Storch-Hagen,
            B.
#submission submitted to the EMBL Data Library, January 1993
#description In vivo comparison of zidovudine resistance in blood and CSF
            of HIV-1 infected patients.
#accession S32079
```

```

2
RESULT      2
ENTRY
TITLE      S63735      #type fragment
CONTAINS   pol polyprotein - human immunodeficiency virus type 1
            (isolate RJ9560) (fragment)
            retropepsin (EC 3.4.23.16); RNA-directed DNA polymerase (EC
            2.7.7.49)
ORGANISM   #formal_name human immunodeficiency virus type 1, HIV-1
            #variety isolate RJ9560
DATE       20-Jul-1996 #sequence_revision 27-Feb-1997 #text_change
            08-Sep-1997
ACCESSIONS S63735; S63707
REFERENCE  #authors Yamaguchi, K.
            #journal Biochim. Biophys. Acta (1995) 1253:136-140
            #title Clinical isolates of HIV-1 contain few pre-existing
            #cross-references MUID:96106422
            #accession S63707
            #molecule_type DNA
            #residues 1-143 ##label YAM
            ##cross-references EMBL:U31389; NID:g961556; PID:g961557
            ##experimental_source isolate RJ9560
REFERENCE  #authors Yamaguchi, K.; Byrn, R.A.
            #journal Biochim. Biophys. Acta (1995) 1253:136-140
            #title Clinical isolates of HIV-1 contain few pre-existing
            #cross-references MUID:96106422
            #accession S63707
            #molecule_type DNA
            #residues 15-113 ##label YAM
            ##cross-references EMBL:U31389
            ##experimental_source isolate RJ9560
GENETICS
#gene      pol
CLASSIFICATION #superfamily pol polyprotein
KEYWORDS   AIDS; aspartic proteinase; hydrolase; immunodeficiency;
            nucleotidyltransferase; polypeptide
FEATURE    15-113
            #product retropepsin #status predicted #label RTP
SUMMARY    #length 143 #checksum 5601
Query Match      85.4%; Score 70; DB 2; Length 143;
Best Local Similarity 100.0%; Pred. No. 1.27e-02;
Matches      8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 132 PKVKQWPL 139
      |||||
Qy 2 PKVKQWPL 9
3
RESULT      3
ENTRY
TITLE      S63745      #type fragment
CONTAINS   pol polyprotein - human immunodeficiency virus type 1
            (isolate RJ14671) (fragment)
            retropepsin (EC 3.4.23.16); RNA-directed DNA polymerase (EC
            2.7.7.49)
ORGANISM   #formal_name human immunodeficiency virus type 1, HIV-1
            #variety isolate RJ14671
DATE       20-Jul-1996 #sequence_revision 27-Feb-1997 #text_change
            08-Sep-1997
ACCESSIONS S63745; S63719
REFERENCE  #authors Yamaguchi, K.
            #submission submitted to the EMBL Data Library, July 1995
            #accession S63745
            #molecule_type DNA
            #residues 1-145 ##label YAM
            ##cross-references EMBL:U31401; NID:g961580; PID:g961581
            ##experimental_source isolate RJ14671
REFERENCE  #authors Yamaguchi, K.; Byrn, R.A.
            #journal Biochim. Biophys. Acta (1995) 1253:136-140

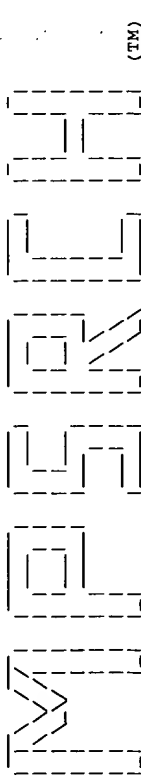
```

```

#title      Clinical isolates of HIV-1 contain few pre-existing
            #cross-references MUID:96106422
            #accession S63719
            #status nucleic acid sequence not shown
            #molecule_type DNA
            #residues 19-74, 'R', 76-117 ##label YAM
            ##cross-references EMBL:U31401
            ##experimental_source isolate RJ14671
CLASSIFICATION #superfamily pol polyprotein
KEYWORDS   AIDS; aspartic proteinase; hydrolase; immunodeficiency;
            nucleotidyltransferase; polypeptide
FEATURE    19-117
            #product retropepsin #status predicted #label RTP
SUMMARY    #length 145 #checksum 315
Query Match      85.4%; Score 70; DB 2; Length 145;
Best Local Similarity 100.0%; Pred. No. 1.27e-02;
Matches      8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 136 PKVKQWPL 143
      |||||
Qy 2 PKVKQWPL 9
4
RESULT      4
ENTRY
TITLE      S63734      #type fragment
CONTAINS   pol polyprotein - human immunodeficiency virus type 1
            (isolate RJ9532) (fragment)
            retropepsin (EC 3.4.23.16); RNA-directed DNA polymerase (EC
            2.7.7.49)
ORGANISM   #formal_name human immunodeficiency virus type 1, HIV-1
            #variety isolate RJ9532
DATE       20-Jul-1996 #sequence_revision 27-Feb-1997 #text_change
            08-Sep-1997
ACCESSIONS S63734; S63706
REFERENCE  #authors Yamaguchi, K.
            #submission submitted to the EMBL Data Library, July 1995
            #accession S63734
            #molecule_type DNA
            #residues 1-149 ##label YAM
            ##cross-references EMBL:U31388; NID:g961554; PID:g961555
            ##experimental_source isolate RJ9532
REFERENCE  #authors Yamaguchi, K.; Byrn, R.A.
            #journal Biochim. Biophys. Acta (1995) 1253:136-140
            #title Clinical isolates of HIV-1 contain few pre-existing
            #cross-references MUID:96106422
            #accession S63706
            #status nucleic acid sequence not shown
            #molecule_type DNA
            #residues 21-119 ##label YAM
            ##cross-references EMBL:U31388
            ##experimental_source isolate RJ9532
GENETICS
#gene      pol
CLASSIFICATION #superfamily pol polyprotein
KEYWORDS   AIDS; aspartic proteinase; hydrolase; immunodeficiency;
            nucleotidyltransferase; polypeptide
FEATURE    21-119
            #product retropepsin #status predicted #label RTP
SUMMARY    #length 149 #checksum 9762
Query Match      85.4%; Score 70; DB 2; Length 149;
Best Local Similarity 100.0%; Pred. No. 1.27e-02;
Matches      8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 138 PKVKQWPL 145
      |||||
Qy 2 PKVKQWPL 9

```

\*\*\*\*\*



\*\*\*\*\*

Release 3.1A John F. Collins, BioComputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Fri Apr 14 23:06:45 2000; MasPar time 3.25 Seconds  
Tabular output not generated. 110.964 Million cell updates/sec

Title: >US-08-452-843-2  
Description: (1-9) from US08452843.pep  
Perfect Score: 82  
Sequence: 1 YPKVKQWPL 9

Scoring table: PAM 150  
Gap 15

Searched: 122810 seqs, 40068593 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: pir62  
1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 24.504; Variance 36.984; scale 0.663

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	70	85.4	141	2	S63750 pol polyprotein - hum	1.27e-02
2	70	85.4	143	2	S63735 pol polyprotein - hum	1.27e-02
3	70	85.4	145	2	S63745 pol polyprotein - hum	1.27e-02
4	70	85.4	149	2	S63734 pol polyprotein - hum	1.27e-02
5	70	85.4	160	2	S63753 pol polyprotein - hum	1.27e-02
6	70	85.4	162	2	S63732 pol polyprotein - hum	1.27e-02
7	70	85.4	219	2	S32128 RNA-directed DNA poly	1.27e-02
8	70	85.4	219	2	S32079 RNA-directed DNA poly	1.27e-02
9	70	85.4	219	2	S32152 RNA-directed DNA poly	1.27e-02
10	70	85.4	219	2	S32159 RNA-directed DNA poly	1.27e-02
11	70	85.4	219	2	S32160 RNA-directed DNA poly	1.27e-02
12	70	85.4	219	2	S32092 RNA-directed DNA poly	1.27e-02
13	70	85.4	219	2	S32094 RNA-directed DNA poly	1.27e-02
14	70	85.4	219	2	S32157 RNA-directed DNA poly	1.27e-02
15	70	85.4	219	2	S32078 RNA-directed DNA poly	1.27e-02
16	70	85.4	219	2	S32093 RNA-directed DNA poly	1.27e-02
17	70	85.4	219	2	S32047 RNA-directed DNA poly	1.27e-02
18	70	85.4	219	2	S32049 RNA-directed DNA poly	1.27e-02
19	70	85.4	219	2	S32048 RNA-directed DNA poly	1.27e-02
20	70	85.4	219	2	S32119 RNA-directed DNA poly	1.27e-02
21	70	85.4	219	2	S32072 RNA-directed DNA poly	1.27e-02
22	70	85.4	219	2	S32089 RNA-directed DNA poly	1.27e-02
23	70	85.4	219	2	S32051 RNA-directed DNA poly	1.27e-02

24	70	85.4	219	2	S32120 RNA-directed DNA poly	1.27e-02
25	70	85.4	219	2	S32073 RNA-directed DNA poly	1.27e-02
26	70	85.4	219	2	S32129 RNA-directed DNA poly	1.27e-02
27	70	85.4	219	2	S32098 RNA-directed DNA poly	1.27e-02
28	70	85.4	219	2	S32066 RNA-directed DNA poly	1.27e-02
29	70	85.4	219	2	S32140 RNA-directed DNA poly	1.27e-02
30	70	85.4	219	2	S32069 RNA-directed DNA poly	1.27e-02
31	70	85.4	219	2	S32133 RNA-directed DNA poly	1.27e-02
32	70	85.4	219	2	S32118 RNA-directed DNA poly	1.27e-02
33	70	85.4	219	2	S32132 RNA-directed DNA poly	1.27e-02
34	70	85.4	219	2	S32080 RNA-directed DNA poly	1.27e-02
35	70	85.4	219	2	S32070 RNA-directed DNA poly	1.27e-02
36	70	85.4	219	2	S32074 RNA-directed DNA poly	1.27e-02
37	70	85.4	219	2	S32096 RNA-directed DNA poly	1.27e-02
38	70	85.4	219	2	S32122 RNA-directed DNA poly	1.27e-02
39	70	85.4	559	2	B47175 reverse transcriptase	1.27e-02
40	70	85.4	559	2	A47175 reverse transcriptase	1.27e-02
41	70	85.4	912	2	S33980 pol polyprotein - hum	1.27e-02
42	70	85.4	1002	1	GNLJND pol polyprotein - hum	1.27e-02
43	70	85.4	1003	1	GNVWLV pol polyprotein - hum	1.27e-02
44	70	85.4	1003	1	B44001 pol polyprotein - hum	1.27e-02
45	70	85.4	1003	1	GNVWAZ pol polyprotein - hum	1.27e-02

ALIGNMENTS

RESULT 1  
ENTRY S63750 #type fragment  
TITLE pol polyprotein - human immunodeficiency virus type 1  
(isolate A012G691-2) (fragment)  
CONTAINS retropepsin (EC 3.4.23.16); RNA-directed DNA polymerase (EC 2.7.7.49)  
ORGANISM #formal name human immunodeficiency virus type 1, HIV-1  
#variety isolate A012G691-2  
DATE 20-Jul-1996 #sequence\_revision 27-Feb-1997 #text\_change 09-May-1997

ACCESSIONS S63750; S63724  
REFERENCE S63731  
#authors Yamaguchi, K.  
#submission submitted to the EMBL Data Library, July 1995  
#accession S63750  
#molecule\_type DNA  
#residues 1-141 #label YAM  
#Cross-references EMBL:U31406  
#experimental\_source isolate A012G691-2  
#authors Yamaguchi, K.; Byrn, R.A.  
#journal Biochim. Biophys. Acta (1995) 1253:136-140  
#title Clinical isolates of HIV-1 contain few pre-existing  
proteinase inhibitor resistance-conferring mutations.  
#cross-references MUID:96106422

#accession S63724  
#status nucleic acid sequence not shown  
#molecule\_type DNA  
#residues 15-113 #label YAM  
#Cross-references EMBL:U31406  
#experimental\_source isolate A012G691-2  
GENETICS  
#gene pol  
#superfamily pol polyprotein  
CLASSIFICATION AIDS; aspartic proteinase; hydrolase; immunodeficiency;  
KEYWORDS nucleotidyltransferase; polypeptide  
FEATURE  
15-113 #product retropepsin #status predicted #label RTP  
SUMMARY  
#length 141 #checksum 1963

Query Match 85.4%; Score 70; DB 2; Length 141;  
Best Local Similarity 100.0%; Pred. No. 1.27e-02;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Db 132 PKVKQWPL 139  
|||||||  
QY 2 PKVKQWPL 9

PD 15-NOV-1990. F00312.  
 PF 02-MAY-1990; FR-005914.  
 PR 03-MAY-1989; FR-005914.  
 PA (INRM ) INSERM INST NAT SANTE.  
 PI Baire-Sinnoussi F, Chermann JC, Devaux C, Rey F, Sire J;  
 PI Spire B; 361470/48.  
 DR WPI; 90-361470/48.  
 DR N-PSDB; Q06635.  
 PT New HIV-NDK retrovirus and protein component - used in vaccines  
 PT against immuno-deficiency disorders and in raising Mabs for  
 PT retro-virus detection in vivo.  
 PS Disclosure: Fig 2; 37pp; French.  
 CC The HIV NDK virus was isolated from peripheral blood lymphocytes of  
 CC an AIDS patient. A genomic library was prepd. from DNA extracted  
 CC from CEM cells infected with the virus. The library was screened  
 CC with a pBT1 probe corresp. to a fragment from HIV 1. The virus is  
 CC more cytopathic than other strains and is not inhibited by OKT4A.  
 CC It has been deposited as CNCM I-857. The sequence can be used to  
 CC express proteins useful for diagnosing the presence of NDK and  
 CC related viruses and in vaccines against immunodeficiency diseases.  
 CC See also R09301-5.  
 SQ Sequence 982 AA;

Query Match 85.4%; Score 70; DB 1; Length 982;  
 Best Local Similarity 100.0%; Pred. No. 1.06e+00; Indels 0; Gaps 0;  
 Matches 8; Conservative 0; Mismatches 0;

Db 153 PKVKQWPL 160  
 QY 2 PKVKQWPL 9  
 |||||

Search completed: Fri Apr 14 23:06:26 2000  
 Job time : 40 secs.

Best Local Similarity 100.0%; Pred. No. 1.06e+00; Mismatches 0; Indels 0; Gaps 0;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 174 PKVKQWPL 181  
|||||||  
QY 2 PKVKQWPL 9

## RESULT 12

ID W52186 standard; Protein; 543 AA.  
AC W52186;  
DT 23-JUN-1998 (first entry)  
DE POL region of p41gag fusion protein in plasmid pGAG41-10.  
KW p41gag gene; enhanced promoter; gene expression; cytomegalovirus;  
KW HIV; AIDS.  
OS Human Immunodeficiency virus.  
FH Location/Qualifiers  
FT Region 523..543  
FT /note= "the nucleotides encoding this region are  
not provided in the specification"

US568688-A.

PD 18-NOV-1997;  
PF 10-AUG-1994; 288336.  
PR 24-DEC-1987; US-138894.  
PR 31-OCT-1984; US-667501.  
PR 30-JAN-1985; US-696534.  
PR 06-SEP-1985; US-773447.  
PR 17-AUG-1992; US-931191.  
PR 28-JUN-1993; US-083391.  
PR 17-AUG-1993; US-107377.  
PR 10-AUG-1994; US-288336.  
PA (CHIR ) CHIRON CORP.  
PI Chapman BS, Dina D, Haigwood NL, Luciw PA, Rosenberg S,  
PI Thayer RM;  
DR WPI; 98-007982/01.  
DR N-PSDB; V04735.  
PT Enhanced promoter for gene expression - comprising cytomegalovirus  
PT Immediate early promoter plus intron  
PS Example 3; Fig 9A-C; 99pp; English.  
CC This is the POL region of the p41gag fusion protein cloned in pGAG41-10.  
CC This is used in the construction of a vector for expression of a  
CC polypeptide in a mammalian cell, comprising a polypeptide coding sequence  
CC operably linked downstream of an enhanced promoter. The enhanced promoter  
CC comprises the human cytomegalovirus immediate early region (HCMV IE1)  
CC promoter and the first intron proximate to the 3' end of the HCMV IE1  
CC promoter. The polypeptide can be any of the HIV recombinant polypeptides  
CC and especially HIV gp120. Expression of HIV gp120 by COS 7 cells  
CC transfected with pcMV6a containing the gp120 coding region, where pcMV6a  
CC is a vector containing the above enhanced promoter, is increased by a  
CC factor of 50-100 compared with the use of a vector containing the SV40  
CC early promoter.  
SQ Sequence 543 AA;

Query Match 85.4%; Score 70; DB 1; Length 543;  
Best Local Similarity 100.0%; Pred. No. 1.06e+00;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 174 PKVKQWPL 181  
|||||||  
QY 2 PKVKQWPL 9

## RESULT 13

ID R08057 standard; protein; 912 AA.  
AC R08057;  
DT 18-JAN-1991 (first entry)  
DE HIV-1 pol protein of HIVMN.  
KW HIV diagnosis; HIV-pol; vaccine; HIVMN;  
KW protein processing; reverse transcriptase; RNase; integrase.  
OS Human immunodeficiency virus - 1.  
PN WO9010230-A.  
PD 07-SEP-1990.  
PF 23-FEB-1990; CA0062.

PR 18-APR-1989; GB-008725.  
PA (UYOT-) UNIV OF OTTAWA.  
PI Kang CY;  
DR WPI; 90-290460/38.

PT Improved polypeptide reagent for HIV diagnosis and vaccine -  
PT comprises portions of all 4 enzymes encoded by HIV-pol gene  
PS Disclosure; Page 11-23; 37pp; English.  
CC Several strains of HIV-1 were cloned and the corresponding amino  
CC acid sequence derived from the determined DNA sequences.  
CC An improved polypeptide reagent comprises portions of all of the  
CC 4 enzymes, and is used in a diagnostic test for HIV infection.  
CC The peptide is also used in vaccines.  
CC See also R08053-83.  
SQ Sequence 912 AA;

Query Match 85.4%; Score 70; DB 1; Length 912;  
Best Local Similarity 100.0%; Pred. No. 1.06e+00;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 83 PKVKQWPL 90  
|||||||  
QY 2 PKVKQWPL 9

## RESULT 14

ID R08053 standard; protein; 912 AA.  
AC R08053;  
DT 18-JAN-1991 (first entry)  
DE ACNPV-HIVK-pol protein of HIVHXB2 virus.  
KW HIV diagnosis; ACNPV-HIVK-pol; vaccine; HIVHXB2;  
KW protein processing; reverse transcriptase; RNase; integrase.  
OS Human immunodeficiency virus - 1.  
PN WO9010230-A.  
PD 07-SEP-1990.  
PF 23-FEB-1990; CA0062.  
PR 18-APR-1989; GB-008725.  
PA (UYOT-) UNIV OF OTTAWA.  
PI Kang CY;  
DR WPI; 90-290460/38.

DR N-PSDB; Q05979.  
PT Improved polypeptide reagent for HIV diagnosis and vaccine -  
PT comprises portions of all 4 enzymes encoded by HIV-pol gene  
PS Disclosure; Page 11-23; 37pp; English.  
CC Recombinant ACNPV-HIVK-pol omits NH2-terminal sequences encoding the  
CC proteolytic active site of the HIV-pol protease. (Compare with ACNPV-  
CC HIVHXB2 (Q06644) comprising the whole DNA sequence of the HIV-pol  
CC gene). When this sequence is expressed the resulting gene product  
CC is not "processed", i.e. the 95 kD protein, comprising HIV-pol  
CC reverse transcriptase, HIV-pol RNase H and HIV-pol integrase,  
CC remains intact.  
CC An improved polypeptide reagent comprises portions of all of the  
CC 4 enzymes, and is used in a diagnostic test for HIV infection.  
CC The peptide is also used in vaccines.  
CC See also R08053-83.  
SQ Sequence 912 AA;

Query Match 85.4%; Score 70; DB 1; Length 912;  
Best Local Similarity 100.0%; Pred. No. 1.06e+00;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 83 PKVKQWPL 90  
|||||||  
QY 2 PKVKQWPL 9

## RESULT 15

ID R09301 standard; protein; 982 AA.  
AC R09301;  
DT 27-FEB-1991 (first entry)  
DE Sequence deduced from pol gene of HIV 1-NDK.  
KW Human immunodeficiency virus; AIDS.  
OS HIV 1-NDK.  
PN WO9013630-A.



CC vectors.  
SQ Sequence 280 AA;

Query Match 85.4%; Score 70; DB 1; Length 280;  
Best Local Similarity 100.0%; Pred. No. 1.06e+00;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 247 PKVKQWPL 254  
|||||||  
Qy 2 PKVKQWPL 9

RESULT 9

ID W1363 standard; Protein; 280 AA.  
AC W1363;  
DT 15-FEB-1999 (first entry)  
DE HIV POL/NEF epitopes.  
KW Vector; gene therapy; vaccine; ALVAC; K3L; E3L; translation factor;  
OS VCP1433; HIV; POL; NEF; epitope.  
PN Human immunodeficiency virus type 1.  
PD 17-SEP-1998.  
PF 13-FEB-1998; U02669.  
PR 12-MAR-1997; US-816155.  
PA (VIRO-) VIROGENETICS CORP.  
PI Cox WI, Gettig RR, Martinez H, Paoletti E, Pincus SE,  
PI Tartaglia J;  
DR WPI: 98-520820/44.  
DR N-PSDB: V60251.  
PT Enhancing expression of nucleic acids in cells - by using modified  
PT vectors which comprise nucleic acid and also nucleic acid encoding  
PT transcription factor and optionally translation factor  
PS Example 2; Fig 6; 102pp; English.  
CC This polypeptide comprises the POL and NEF epitopes of HIV-1.  
CC It is encoded by vcp1433 (see V60251). In vcp1433, an HIV pol/nef  
CC 'string of beads' cassette is placed under control of the  
CC vaccinia H6 promoter. pMPC6H6K3E3, containing a vaccinia  
CC H6/K3L expression cassette and vaccinia E3L gene with endogenous  
CC promoter flanked by the ALVAC C6 insertion site sequences, was used  
CC in recombination with vcp1433 to obtain vcp1452 (see V60252-53).  
CC K3L and E3L are vaccinia virus translation factors. New vectors  
CC are provided for enhanced expression of at least 1 first nucleic  
CC acid molecule (NAM) in a cell having a particular phenotype. The  
CC vector (e.g. NVVAC or ALVAC) is modified to comprise the first NAM  
CC and at least 1 second NAM encoding a transcription factor (TF), or  
CC a TF and a translation factor such as K3L and E3L, where there is  
CC co-temporal expression of the first and second NAMs with respect to  
CC the phenotype of the cell, and where expression of the second NAM  
CC enhances expression of the first NAM by enhancing transcription or  
CC transcription and translation. Also claimed is a method for  
CC increasing expression of at least 1 first NAM by such a vector.  
CC The vectors can be used for increasing expression of e.g. an  
CC epitope of interest, a biological response modulator, a growth  
CC factor, a recognition sequence, a therapeutic gene or a fusion  
CC protein.  
SQ Sequence 280 AA;

Query Match 85.4%; Score 70; DB 1; Length 280;  
Best Local Similarity 100.0%; Pred. No. 1.06e+00;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 247 PKVKQWPL 254  
|||||||  
Qy 2 PKVKQWPL 9

RESULT 10

ID R29709 standard; Protein; 522 AA.  
AC R29709;  
DT 26-APR-1993 (first entry)  
DE p41 gag protein from HTLV.  
KW express proteins; lymphadenopathy syndrome; AIDS; HIV; HTLV;  
KW serological immunoassays; antibodies to HTLV; monoclonal antibodies;

KW probes; ss.  
OS Human T cell lymphotropic retrovirus.  
PN EP-518443-A.  
PD 16-DEC-1992.  
PF 30-OCT-1985; 201711.  
PR 31-OCT-1984; US-667501.  
PR 30-JAN-1985; US-696534.  
PR 06-SEP-1985; US-773447.  
PA (CHIR-) CHIRON CORP.  
PI Barr PJ, Dina D, George-Nascimento C, Hallewell R;  
PI Luciw PA, Parkes D, Pescador RS, Steimer K, Truett M;  
DR WPI: 92-417329/51.  
DR N-PSDB: Q31938.  
PT Recombinant DNA construct including replication system recognised  
PT by unicellular microorganism - used to form recombinant proteins  
PT for diagnosing AIDS and lymphadenopathy syndrome  
PS Example 11; Fig 5; 32pp; English.  
CC This sequence was decoded from the p41 gag gene from HTLV DNA.  
CC Proteins associated with lymphadenopathy syndrome and/or AIDS may  
CC be used in serological immunoassays to detect antibodies to HTLV.  
CC The polypeptides can be used alone or in fusion constructs to  
CC produce antisera or monoclonal antibodies which may be used for  
CC therapy or diagnosis.  
SQ Sequence 522 AA;

Query Match 85.4%; Score 70; DB 1; Length 522;  
Best Local Similarity 100.0%; Pred. No. 1.06e+00;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 174 PKVKQWPL 181  
|||||||  
Qy 2 PKVKQWPL 9

RESULT 11

ID P61510 standard; Protein; 543 AA.  
AC P61510;  
DT 13-SEP-1991 (first entry)  
DE Sequence of pol protein encoded by ARV-2 cDNA cloned in pGAG41-10  
DE for producing the fusion protein p41 gag.  
KW LAV; HIV; ARV; HTLV; vaccine; AIDS; Immunoassay; diagnosis;  
KW lymphadenopathy syndrome.  
OS Human T-cell lymphotropic virus III.  
PN EP-181150-A.  
PD 14-MAY-1986.  
PF 30-OCT-1985; 307860.  
PR 31-OCT-1984; US-667501.  
PR 30-JAN-1985; US-696534.  
PR 06-SEP-1985; US-773447.  
PA (CHIR-) CHIRON CORP.  
PI Luciw PA, Dina D, Steimer K, Pescador RS, George-Nascimento C,  
PI Parkes D, Hallewell R, Barr PJ, Truett M;  
DR WPI: 86-126568/20.  
DR N-PSDB: N60142.  
PT New recombinant human T-cell lymphotropic retro virus proteins -  
PT useful in diagnostic immunoassays for antibodies in humans, and  
PT in prodn. of monoclonal antibodies, as vaccines etc.  
PS Disclosure; Fig 5; 67pp; English.  
CC The inventors claim a DNA construct contg. a DNA sequence  
CC substantially as set forth in N60141, N60142, N60143, N60144, which  
CC are each derived from AIDS-associated retroviruses. For the purposes  
CC of this application, HTLV-III, LAV and ARV are generically referred  
CC to as human T-cell lymphotropic retrovirus (HTLV). The following  
CC recombinant polypeptides are also claimed: (a) ARV-2 p16 gag;  
CC (b) ARV-2 p25 gag; (c) ARV-2 env; (d) ARV-2 p31 pol. pGAG was  
CC constructed from plasmid pGAG25-10 by inserting an SphI-HpaI  
CC fragment from the ARV-2 genome containing the sequences from the  
CC C-terminal p16 gag portion of the p53 gag precursor polyprotein and  
CC part of the p25 gag protein between the SphI and BamHI sites of  
CC pGAG25-10.  
SQ Sequence 543 AA;

Query Match 85.4%; Score 70; DB 1; Length 543;

```
SQ Sequence 25 AA;
Query Match 85.4%; Score 70; DB 1; Length 25;
Best Local Similarity 100.0%; Pred. No. 1.06e+00;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 15 PKVKQWPL 22
|||||
QY 2 PKVKQWPL 9

RESULT 6
ID R94761 standard; Protein; 49 AA.
AC R94761;
DT 11-JUN-1996 (first entry)
DE CTL epitopes polI derived from pCPOLn5A.
KW Canyapox; CPV; ALVAC; attenuated; therapy; prevention; rabies;
KW vector; vaccine; antibody; CTL1; CTL2.
OS Synthetic.
PN W09527507-A1.
PD 19-OCT-1995.
PF 06-APR-1995; U03989.
PR 06-APR-1994; US-223842.
PR 05-APR-1995; US-417210.
PA (VIRO-) VIROGENETICS CORP.
PI Cox WI, Paoletti E, Tartaglia J;
DR WPI; 95-366231/47.
DR N-PSDB; T04705.
PT Virulence-attenuated virus encoding an immunodeficiency virus
PT epitope - based on Copenhagen strain of vaccinia virus, used in
PT prevention and treatment of diseases, e.g. vaccination against HIV
PS Example 16; Fig 18; 208pp; English.
CC This sequence is a pC5POLn5A-derived CTL polI epitope. pC5POLn5A is
CC a plasmid contg. attenuated virus ALVAC recombinant expressing 3 CTL
CC pol epitopes, HIV1 gag (+pro) (IIIB) and gp120 (MN) and transmembrane
CC region. ALVAC-based recombinant viruses expressing extrinsic immunogens
CC are efficacious as vaccine vectors. Attenuated recombinant viruses such
CC as ALVAC or NYVAC can be engineered to comprise exogenous DNA in a non-
CC essential region of their genome, the exogenous DNA encodes at least one
CC immunodeficiency virus epitope. Such attenuated viruses (as above) and
CC derived antigens and antibodies are used in the prevention, therapy
CC and diagnosis of diseases. DNA from the recombinant viruses can be used
CC as probes or for generating primers or for immunisation. Attenuated,
CC recombinant viruses have enhanced safety making them safer for use in
CC vaccines.
SQ Sequence 49 AA;

Query Match 85.4%; Score 70; DB 1; Length 49;
Best Local Similarity 100.0%; Pred. No. 1.06e+00;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 16 PKVKQWPL 23
|||||
QY 2 PKVKQWPL 9

RESULT 7
ID W53109 standard; Protein; 163 AA.
AC W53109;
DT 25-JUN-1998 (first entry)
DE Protein 2 contained in a complete ARV-2 nucleotide sequence.
KW ARV-2; enhanced promoter; gene expression; cytomegalovirus;
KW HIV; AIDS.
OS Human immunodeficiency virus type 1.
PN US5688688-A.
PD 18-NOV-1997.
PF 10-AUG-1994; 288336.
PR 24-DEC-1987; US-138894.
PR 31-OCT-1984; US-567501.
PR 30-JAN-1985; US-696534.
PR 06-SEP-1985; US-773447.
PR 17-AUG-1992; US-931191.
PR 28-JUN-1993; US-083391.

17-AUG-1993; US-107377.
10-AUG-1994; US-288336.
(CHIR ) CHIRON CORP.
Chapman BS, Dina D, Haigwood NL, Luciw PA, Rosenberg S,
Thayer RM;
WPI; 98-007982/01.
N-PSDB; V04733.
Enhanced promoter for gene expression - comprising cytomegalovirus
Immediate early promoter plus intron
Example 1; Fig 4C-P; 99pp; English.
This sequence represents a protein of unspecified function contained in a
complete nucleotide sequence of ARV-2 derived from partial sequences of
several ARV clones. The invention provides a method for construction of a
vector for expression of a polypeptide in a mammalian cell, comprising a
polypeptide coding sequence operably linked downstream of an enhanced
promoter. The enhanced promoter comprises the human cytomegalovirus
immediate early region (HCMV IE1) promoter and the first intron proximate
to the 3' end of the HCMV IE1 promoter. The polypeptide can be any of the
HIV recombinant polypeptides and especially HIV gp120. Expression of HIV
gp120 by COS 7 cells transfected with pCMV6a containing the gp120 coding
region, where pCMV6a is a vector containing the above enhanced promoter,
is increased by a factor of 50-100 compared with the use of a vector
containing the SV40 early promoter.
SQ Sequence 163 AA;

Query Match 85.4%; Score 70; DB 1; Length 163;
Best Local Similarity 100.0%; Pred. No. 1.06e+00;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 83 PKVKQWPL 90
|||||
QY 2 PKVKQWPL 9

RESULT 8
ID W71642 standard; Protein; 280 AA.
AC W71642;
DT 18-JAN-1999 (first entry)
DE HIV POL/NEF epitopes.
KW Vector; gene therapy; vaccine; ALVAC; translation factor; K3L; E3L;
KW VCP1433; HIV; pol; nef; epitope.
OS Human immunodeficiency virus type 1.
PN W09840500-A1.
PD 17-SEP-1998.
PF 25-FEB-1998; U03710.
PR 12-MAR-1997; US-815809.
PA (UYAR-) UNIV ARIZONA STATE.
PA (VIRO-) VIROGENETICS CORP.
PI Cox WI, Gettig RR, Goebel SJ, Jacobs BL, Paoletti E,
Pincus SE, Tartaglia J;
DR WPI; 98-520819/44.
DR N-PSDB; V58243.
PT Enhancing expression of nucleic acids in cells - by using modified
PT vectors which comprise the nucleic acid and also nucleic acid
PT encoding a translation factor
Example 1; Fig 4A-C; 90pp; English.
This polypeptide comprises the POL and NEF epitopes of HIV-1.
It is encoded by VCP1433 (see V58243). In VCP1433, an HIV pol/nef
'string of beads' cassette is placed under control of the
vaccinia H6 promoter. pMPC6H6K3E3, containing a vaccinia
H6/K3L expression cassette and vaccinia E3L gene with endogenous
promoter flanked by the ALVAC C6 insertion site sequences, was used
in recombinant with VCP1433 to obtain VCP1452 (see V58244-45).
CC K3L and E3L are vaccinia virus translation factors. Novel vectors
of the invention, such as ALVAC vectors, include K3L and/or E3L and
are used for enhancing expression of gene products that they
encode. The translation factors can effect inhibition of
eIF-2alpha phosphorylation or inhibition of protein kinase PKR
phosphorylation or otherwise sequester double stranded (ds) RNA,
increasing the effective concentration of ds RNA. The up-regulation
of foreign gene expression can have a profound effect on the
induction of a therapeutic or immunological response in a host
administered or inoculated with recombinants derived from these new
```

DE HIV pol 185-193 cytotoxic T lymphocyte epitope.  
 KW HIV pol 185-193; cytotoxic T; CTL; epitope; helper T; HTL; cell;  
 KW lymphocyte; viruses; parasites; tumours; antigens; treatment;  
 KW disease prevention.  
 OS Human immunodeficiency virus.  
 PN WO9522317-A1.  
 PD 24-AUG-1995.  
 PF 16-FEB-1995; U02121.  
 PR 16-FEB-1994; US-197484.  
 PA (CYTE-) CYTEL CORP.  
 PI Ceut RW, Grey H, Sette AD, Vitiello MA;  
 DR WPI: 95-302545/39.  
 PT Compn. inducing cytotoxic T lymphocyte response to pref. viral,  
 PT bacterial, parasitic or tumour antigens - useful in the treatment  
 PT and prevention of diseases associated with the antigen e.g.  
 PT hepatitis B  
 PS Disclosure: Page 17: 109pp; English.  
 CC A compn. which induces a cytotoxic T lymphocyte (CTL) response to  
 CC an antigen (Ag) in a mammal comprises, a CTL Ag response inducing  
 CC peptide (i.e. R78824-R78853) and a lipid conjugated helper T cell  
 CC inducing peptide. The compn. induces a CTL response to bacterial,  
 CC viral or tumour Ags, and is therefore useful in the treatment and  
 CC prevention of diseases associated with the Ag.  
 CC Sequence 9 AA;  
 SQ

Query Match 85.4%; Score 70; DB 1; Length 9;  
 Best Local Similarity 100.0%; Pred. No. 1.06e+00;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 2 PKVKQWPL 9  
 |||||  
 QY 2 PKVKQWPL 9

## RESULT 3

ID R70601 standard; Peptide; 9 AA.  
 AC R70601;  
 DT 14-FEB-1996 (first entry)  
 DE HIV(B35)POL-9, human immunodeficiency virus epitope.  
 KW HLA; human lymphocyte antigen; HIV; human immunodeficiency virus;  
 KW binding peptide; induce killer cell; prevention; treatment; AIDS;  
 KW autoimmune disease syndrome; vaccine.  
 OS Human immunodeficiency virus.  
 PN WO9511255-A1.  
 PD 27-APR-1995.  
 PF 19-OCT-1994; J01756.  
 PR 19-OCT-1993; JP-261302.  
 PA (AJIN) AJINOMOTO CO INC.  
 PA (AJIN) AJINOMOTO KK.  
 PI Miwa K, Takiguchi M;  
 DR WPI: 95-170188/22.  
 PT HLA-binding peptide fragments from HIV proteins - induce killer  
 PT cells which target HIV-infected cells and can be incorporated into  
 PT anti-HIV vaccines  
 PS Example 1; Page 10; 61pp; Japanese.  
 CC R70601 is a peptide fragment derived from an HIV (Human Immunodeficiency  
 CC Virus) protein and is capable of binding to a human lymphocyte antigen.  
 CC The peptide can induce killer cells which target HIV-infected cells.  
 CC It is also useful in the prevention and treatment of HIV and AIDS.  
 CC Anti-HIV vaccines may incorporate the peptides, or may incorporate a  
 CC vector (such as vaccinia or BCG) contg. DNA encoding the peptides.  
 CC Sequence 9 AA;  
 SQ

Query Match 85.4%; Score 70; DB 1; Length 9;  
 Best Local Similarity 100.0%; Pred. No. 1.06e+00;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 2 PKVKQWPL 9  
 |||||  
 QY 2 PKVKQWPL 9

## RESULT 4

ID W32893 standard; peptide; 25 AA.  
 AC W32893;  
 DT 16-JAN-1998 (first entry)  
 DE HIV pol protein epitope 40.  
 KW Hydrophilic; antigenic determinant; HIV; envelope; glycoprotein;  
 KW env; gp; recognition; B lymphocyte; type specific; antibody;  
 KW vaccine; protection; immune response; infection; neutralisation;  
 KW epitope.  
 OS Human immunodeficiency virus.  
 PN WO9714436-A1.  
 PD 24-APR-1997.  
 PF 18-OCT-1996; U16911.  
 PR 09-FEB-1996; US-599266.  
 PR 20-OCT-1995; US-546515.  
 PA (UYDU-) UNIV DUKE.  
 PI Haynes BF, Parker TJ;  
 DR WPI: 97-244862/22.  
 PT Synthetic human immunodeficiency virus vaccine - comprising  
 PT hydrophilic peptide corresponding to at least 1 antigenic  
 PT determinant of envelope glycoprotein recognised by B lymphocytes  
 PS Disclosure: Page 27; 104pp; English.  
 CC An essentially pure hydrophilic peptide, comprising at least 1  
 CC antigenic determinant of human immunodeficiency virus (HIV)  
 CC envelope (env) glycoprotein (gp) recognised by B lymphocytes,  
 CC when covalently linked to a carrier molecule, i.e. the present  
 CC sequence, induces the production of high titres of protective, type  
 CC specific anti-HIV antibodies (Ab) in a mammal. The peptide can be  
 CC used in vaccines for producing a protective immune response to HIV  
 CC infection, while a HIV neutralising Ab can be induced in a primate  
 CC by administering a composition comprising HIV env peptides that  
 CC disrupt gp120/gp41 interactions.  
 CC Sequence 25 AA;  
 SQ

Query Match 85.4%; Score 70; DB 1; Length 25;  
 Best Local Similarity 100.0%; Pred. No. 1.06e+00;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 15 PKVKQWPL 22  
 |||||  
 QY 2 PKVKQWPL 9

## RESULT 5

ID R68755 standard; peptide; 25 AA.  
 AC R68755;  
 DT 23-AUG-1995 (first entry)  
 DE Cytotoxic T lymphocyte epitope 12 derived from pol protein.  
 KW cytotoxic T lymphocyte; epitope; antigen; pathogenic; nef; gag; pol;  
 KW env; gp120; gp41; HIV; cell-mediated immunity;  
 KW Human immunodeficiency virus; class I restricted.  
 OS Human immunodeficiency virus.  
 PN WO9428871-A.  
 PD 22-DEC-1994.  
 PF 07-JUN-1994; U06394.  
 PR 07-JUN-1993; US-072718.  
 PA (ENDO-) ENDOCON INC.  
 PA Leonard RJ;  
 DR WPI: 95-036067/05.  
 PT Implant for sustained release of pathogen-associated antigen -  
 PT forming chronic inflammatory site producing cytotoxic  
 PT T-lymphocytes lysing infected cells, esp. for treating AIDS  
 PS Disclosure: Page 11; 35pp; English.  
 CC R68744-805 are cytotoxic T lymphocyte (CTL) class I and II restricted  
 CC epitopes derived from human immunodeficiency virus proteins. R68755  
 CC corresponds to amino acid residues 172-196 of the pol protein. These  
 CC antigens are examples of peptides that can be used with an immunogenic  
 CC implant. The implant is associated with an antigen associated with a  
 CC pathogen and used to form a discrete, localised chronic inflammation  
 CC site which acts as a local 'factory' for prodn. of CTL's which lyse  
 CC cells infected with a specific pathogen. The expanded set of  
 CC pathogen-specific CTL's can eradicate or prevent development of  
 CC infection, and can also be used to treat or arrest the development of  
 CC cancers associated with infection.

\*\*\*\*\*  
M P S R E H  
(TM)  
\*\*\*\*\*

Release 3.1A John F. Collins, BioComputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Fri Apr 14 23:05:46 2000; MasPar time 4.16 Seconds  
Tabular output not generated. 51.185 Million cell updates/sec

Title: >US-08-452-843-2  
Description: (1-9) from US08452843.pep  
Perfect Score: 82  
Sequence: 1 YPKVKQWPL 9

Scoring table: PAM 150  
Gap 15

Searched: 188963 seqs, 23686106 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: a-geneseq36  
1:geneseqp

Statistics: Mean 18.131; Variance 54.308; scale 0.334

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	82	100.0	9	1 R89363	Immunogenic peptide, b	4.13e-02
2	70	85.4	9	1 R78851	HIV pol 185-193 cyto	1.06e+00
3	70	85.4	9	1 R70801	HIV(B35)POL-9, human	1.06e+00
4	70	85.4	25	1 W32893	HIV pol protein epitop	1.06e+00
5	70	85.4	25	1 R68755	Cytotoxic T lymphocyte	1.06e+00
6	70	85.4	49	1 R94761	CTL epitopes pol1 deri	1.06e+00
7	70	85.4	163	1 W53109	Protein 2 contained in	1.06e+00
8	70	85.4	280	1 W71642	HIV POL/NEF epitopes.	1.06e+00
9	70	85.4	280	1 W71363	HIV POL/NEF epitopes.	1.06e+00
10	70	85.4	522	1 R29709	p41 gag protein from h	1.06e+00
11	70	85.4	543	1 P61510	Sequence of pol protei	1.06e+00
12	70	85.4	543	1 W52186	POL region of p41gag f	1.06e+00
13	70	85.4	912	1 R08057	HIV-1 pol protein of H	1.06e+00
14	70	85.4	912	1 R08053	ACNPV-HIVK-pol protei	1.06e+00
15	70	85.4	982	1 R09301	Sequence deduced from	1.06e+00
16	70	85.4	1001	1 R12256	HIV-1 strain OYI POL p	1.06e+00
17	70	85.4	1003	1 P61508	Sequence of ARV-2 (9B)	1.06e+00
18	70	85.4	1003	1 R08061	HIV-1 pol protein of H	1.06e+00
19	70	85.4	1003	1 R29705	pol gene decoded from	1.06e+00
20	70	85.4	1003	1 P60420	Sequence of LAV virus	1.06e+00
21	70	85.4	1003	1 P70861	Sequence encoded by LA	1.06e+00
22	70	85.4	1003	1 R08059	HIV-1 pol protein of H	1.06e+00
23	70	85.4	1004	1 R08038	HIV-1 pol protein of H	1.06e+00

24	70	85.4	1010	1 R91823	Human immunodeficiency	1.06e+00
25	70	85.4	1012	1 W90176	HTLV-III pol protein.	1.06e+00
26	70	85.4	1012	1 W89323	HIV-1 pol protein sequ	1.06e+00
27	70	85.4	1012	1 P61507	Sequence of reverse tr	1.06e+00
28	70	85.4	1014	1 W68474	HIV-1 strain YBF30 pol	1.06e+00
29	70	85.4	1015	1 P60347	HTLV-III virus (HIV v1	1.06e+00
30	70	85.4	1015	1 R43875	HTLV-III POL gene prod	1.06e+00
31	70	85.4	1015	1 R43867	HTLV-III POL gene prod	1.06e+00
32	70	85.4	1016	1 R08062	ACNPV-HIVHPol protein	1.06e+00
33	70	85.4	1016	1 R08054	HIV-1 pol protein of H	1.06e+00
34	70	85.4	1016	1 R08063	HIV-1 pol protein of H	1.06e+00
35	70	85.4	1022	1 P81854	Sequence encoded by LA	1.06e+00
36	70	85.4	1491	1 P91048	Transcription sequence	1.06e+00
37	70	85.4	2033	1 R08056	HIV-1 pol protein of H	1.06e+00
38	70	85.4	2033	1 R08055	HIV-1 pol protein of H	1.06e+00
39	67	81.7	1002	1 W72993	HIV isolate LAV.MAL po	2.33e+00
40	67	81.7	1002	1 P81861	Sequence encoded by LA	2.33e+00
41	67	81.7	1003	1 R08060	HIV-1 pol protein of H	2.33e+00
42	67	81.7	1055	1 W13055	HIV-2 provirus-encoded	2.33e+00
43	65	79.3	1009	1 R10275	Simian immunodeficien	3.93e+00
44	64	78.0	1056	1 P80809	Sequence of pol protei	5.10e+00
45	64	78.0	1060	1 W89314	STVmac239 genome pol p	5.10e+00

ALIGNMENTS

RESULT 1  
ID R89363 standard; peptide; 9 AA.  
AC R89363:  
DE 18-SEP-1996 (first entry)  
DE Immunogenic peptide, based on Y1 analog of 1054.05.  
KW Immunogenic peptide; supermotif; HLA molecule; CTL response;  
KW therapeutic; diagnostic; cancer; viral infection; hepatitis B;  
KW hepatitis C.  
OS Synthetic.  
PN WO9603140-A1.  
PD 08-FEB-1996.  
PF 21-JUL-1995; U092334.  
PR 21-JUL-1994; US-378634.  
PR 23-NOV-1994; US-344824.  
PR 30-MAY-1995; US-452843.  
PA (CYTE-) CYTEL CORP.  
PI Sette A, Sidney J;  
PI WPI; 96-116784/12.  
PT Compsn. comprising immunogenic peptide with supermotif allowing more  
PT than one HLA mol. to bind - used to induce CTL response in patient  
PT and for in vivo and ex vivo therapeutic and diagnostic applications  
PS Claim 2; Page 26; 32pp; English.  
CC The sequences given in R89362-82 are immunogenic peptides which were  
CC use in the composition of the invention. The composition comprises  
CC an immunogenic peptide of 9-10 residues with a supermotif which  
CC allows binding of more than one HLA molecule. It pref. comprises  
CC two conserved residues, a first at the 2nd position from the N-  
CC terminal is Pro, and a 2nd at the C-terminal is Met. These peptides  
CC are used to induce a CTL response in a patient. They are also  
CC useful in compositions for in vivo and ex vivo therapeutic and  
CC diagnostic applications, e.g the treatment of cancer and viral  
CC infections, e.g. hepatitis B and C.  
SQ Sequence 9 AA;

Query Match 100.0%; Score 82; DB 1; Length 9;  
Best Local Similarity 100.0%; Pred. No. 4.13e-02;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 YPKVKQWPL 9  
QY 1 YPKVKQWPL 9  
|||||

RESULT 2  
ID R78851 standard; peptide; 9 AA.  
AC R78851:  
DT 27-MAR-1996 (first entry)

Job time : 102 secs.

.....

QY 1 FPFKYAAAF 9

RESULT 12 PRELIMINARY; PRT; 1395 AA.  
ID O43168  
AC O43168;  
DT 01-JUN-1998 (TRENBLrel. 06, Created)  
DT 01-JUN-1998 (TRENBLrel. 06, Last sequence update)  
DT 01-AUG-1998 (TRENBLrel. 07, Last annotation update)  
DE KIA0443.  
GN KIA0443.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=BRAIN;  
RA ISHIKAWA K., NAGASE T., NAKAJIMA D., SEKI N., OHIRA M., MIYAJIMA N.,  
RA TANAKA A., KOTANI H., NOMURA N., OHARA O.;  
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AB007903; BAA23715.1; -;  
SQ SEQUENCE 1395 AA; 156836 MW; A0976EF7 CRC32;

Query Match 72.2%; Score 52; DB 4; Length 1395;  
Best Local Similarity 44.4%; Pred. No. 1.83e+01;  
Matches 4; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 1111 FPFQDPSP 1119

| | | | : :

QY 1 FPFKYAAAF 9

RESULT 13 PRELIMINARY; PRT; 139 AA.  
ID Q23322;  
AC Q23322;  
DT 01-NOV-1996 (TRENBLrel. 01, Created)  
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)  
DT 01-JAN-1999 (TRENBLrel. 09, Last annotation update)  
DE ZC443.2 PROTEIN.  
GN ZC443.2  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Secernentes; Rhabditia; Rhabditida;  
OC Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA BAYNES C.;  
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RX MEDLINE; 94150718.  
RA WILSON R., AINSWORTH R., ANDERSON K., BAYNES C., BERKS M.,  
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,  
RA CRATON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,  
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,  
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,  
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,  
RA PARSONS J., PERCY C., RIFKEN L., ROOPER A., SAUNDERS D., SHOWNKEEN R.,  
RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,  
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,  
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;  
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
elegans.";  
RL Nature 368:32-38(1994).  
DR EMBL: Z75553; CAA99949.1; -;  
SQ SEQUENCE 139 AA; 16819 MW; 1C24D160 CRC32;

Query Match 70.8%; Score 51; DB 5; Length 139;  
Best Local Similarity 55.6%; Pred. No. 2.75e+01;  
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 66 FRPHYSASF 74

| | | | : :

QY 1 FPFKYAAAF 9

RESULT 14 PRELIMINARY; PRT; 209 AA.  
ID O68567;  
AC O68567;  
DT 01-AUG-1998 (TRENBLrel. 07, Created)  
DT 01-AUG-1998 (TRENBLrel. 07, Last sequence update)  
DT 01-AUG-1998 (TRENBLrel. 07, Last annotation update)  
DE RESTRICTION ENDONUCLEASE R.XBAI.  
GN XBAIR.  
OS Xanthomonas campestris.  
OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;  
OC Xanthomonas.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BADRII;  
RA ZHANG B.-H., WILSON G.G.;  
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AF051092; AAC08983.1; -;  
KW Endonuclease.  
SQ SEQUENCE 209 AA; 23834 MW; 1EC63D0B CRC32;

Query Match 70.8%; Score 51; DB 2; Length 209;  
Best Local Similarity 55.6%; Pred. No. 2.75e+01;  
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 62 FLYKYAGSF 70

| : | | | : | |

QY 1 FPFKYAAAF 9

RESULT 15 PRELIMINARY; PRT; 211 AA.  
ID O26053;  
AC O26053;  
DT 01-JAN-1998 (TRENBLrel. 05, Created)  
DT 01-JAN-1998 (TRENBLrel. 05, Last sequence update)  
DT 01-NOV-1998 (TRENBLrel. 08, Last annotation update)  
DE HYPOTHETICAL 24.9 KD PROTEIN.  
GN HP1525.  
OS Helicobacter pylori (Campylobacter pylori).  
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;  
OC Helicobacter.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=26695;  
RX MEDLINE; 97394467.  
RA TOMB J.-F., WHITE O., KERLAVAGE A.R., CLAYTON R.A., SUTTON G.G.,  
RA FLEISCHMANN R.D., KETCHUM K.A., KLENK H.-P., GILL S., DOUGHERTY B.A.,  
RA NELSON K., QUACKENBUSH J., ZHOU L., KIRKNESS E.F., PETERSON S.,  
RA LOFTUS B., RICHARDSON D., DODSON R., KHALAK H.G., GLODEK A.,  
RA MCKENNEY K., FITZGERALD L.M., LEE N., ADAMS M.D., HICKEY E.K.,  
RA BERG D.E., GOCAYNE J.D., UTTERBACK T.R., PETERSON J.D., KELLEY J.M.,  
RA COTTON M.D., WEIDMAN J.M., FUJII C., BOWMAN C., WATTHEY L., WALLIN E.,  
RA HAYES W.S., BORODOVSKY M., KARP P.D., SMITH H.O., FRASER C.M.,  
RA VENTER J.C.;  
RT "The complete genome sequence of the gastric pathogen Helicobacter  
pylori.";  
RL Nature 388:539-547(1997).  
DR EMBL: AE000650; AAO08567.1; -;  
DR TIGR; HP1525; -;  
KW Hypothetical protein.  
SQ SEQUENCE 211 AA; 24866 MW; 6BED7882 CRC32;

Query Match 70.8%; Score 51; DB 2; Length 211;  
Best Local Similarity 55.6%; Pred. No. 2.75e+01;  
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 189 FAFYDVSASF 197

| : | | | : | |

QY 1 FPFKYAAAF 9

Search completed: Fri Apr 14 23:03:49 2000

```

DR EMBL; 223281; CAA80819.1; -.
DR PFAM; PF01473; CW_binding_1; 10.
SQ SEQUENCE 2178 AA; 250134 MW; 0C347C36 CRC32;

Query Match 75.0%; Score 54; DB 2; Length 2178;
Best Local Similarity 66.7%; Pred.No. 7.92e+00;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 1273 FPKYEAFF 1281
   1 FPKYAAAF 9
   |||||
RESULT 9
ID O42181 PRELIMINARY; PRT; 4578 AA.
AC O42181;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
DE PKD1 PROTEIN.
GN PKD1.
OS Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Actinopterygii;
OC Neopterygii; Teleostei; Euteleostei; Acanthopterygii; Percomorpha;
OC Tetraodontiformes; Tetraodontidae; Tetraodontidae; Fugu.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 97449170.
RA SANDFORD R., SGOTTO B., APARICIO S., BRENNER S., VAUDIN M., WILSON R.,
RA CHISSOE S., PEPIN K., BATEMAN A., CHOTHIA C., HUGHES J., HARRIS P.;
RT "Comparative analysis of the polycystic kidney disease 1 (PKD1) gene
RT reveals an integral membrane glycoprotein with multiple evolutionary
RT conserved domains."
RL Hum. Mol. Genet. 6:1483-1489(1997).
RN [2]
RP SEQUENCE FROM N.A.
RA VAUDIN M.;
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA WASHU;;
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA WATERSTON R.;
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF013614; AAB86683.1; -.
DR PFAM; PF00059; Lectin_c; 1.
DR PFAM; PF01463; LRCT; 1.
DR PFAM; PF00801; PKD; 14.
DR PFAM; PF01477; PLAT; 1.
SQ SEQUENCE 4578 AA; 504591 MW; DE8EE954 CRC32;

Query Match 75.0%; Score 54; DB 13; Length 4578;
Best Local Similarity 55.6%; Pred.No. 7.92e+00;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 2951 FPFYVANY 2959
   1 FPFYAAAF 9
   |||||
RESULT 10
ID O86755 PRELIMINARY; PRT; 208 AA.
AC O86755;
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)
DE HYPOTHETICAL 23.3 KD PROTEIN.
GN SC6A9.23C.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.

```

---

```

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA MURPHY L., HARRIS D.;
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA PARKHILL J., BARRELL B.G., RAJANDREAM M.A.;
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RX MEDLINE; 97000351.
RA REDENBACH M., KIESER H.M., DENAPAITE D., EICHNER A., CULLUM J.,
RA KINASHI H., HOPWOOD D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome."
RL Mol. Microbiol. 21:77-96(1996).
DR EMBL; AL031035; CAA19908.1; -.
KW Hypothetical protein.
SQ SEQUENCE 208 AA; 23346 MW; E1B3EC9F CRC32;

Query Match 73.6%; Score 53; DB 2; Length 208;
Best Local Similarity 100.0%; Pred.No. 1.21e-01;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 96 FPFYA 101
   |||||
QY 1 FPFYA 6

RESULT 11
ID Q23370 PRELIMINARY; PRT; 828 AA.
AC Q23370;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-MAY-1999 (TREMBLrel. 10, Last annotation update)
DE ZC518.3 PROTEIN.
GN ZC518.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Secernentes; Rhabditia; Rhabditidae;
OC Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN [1]
RP SEQUENCE FROM N.A.
RA THOMAS K.;
RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans."
RL Nature 368:32-38(1994).
DR EMBL; Z68753; CAA92990.1; -.
DR PFAM; PF00560; LRR; 3.
SQ SEQUENCE 828 AA; 94021 MW; 266AAD6D CRC32;

Query Match 72.2%; Score 52; DB 5; Length 828;
Best Local Similarity 55.6%; Pred.No. 1.83e+01;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 434 FPFNFHATF 442
   |||||

```

RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,  
 RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,  
 RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,  
 RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,  
 RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,  
 RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULTON J.,  
 RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,  
 RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.,  
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
 RT elegans";  
 RT Nature 368:32-38(1994).  
 DR EMBL; Z77665; CAB01223.1; -  
 SQ SEQUENCE 800 AA; 90383 MW; A52285D2 CRC32;

Query Match 76.4%; Score 55; DB 5; Length 800;  
 Best Local Similarity 75.0%; Pred. No. 5.18e+00;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 529 FPFKYTNA 536  
 QY 1 FPFKYAAA 8  
 |||||: 1

RESULT 6  
 ID O16976 PRELIMINARY; PRT; 466 AA.  
 AC O16976;  
 DT 01-JAN-1998 (TRENBLrel. 05, Created)  
 DT 01-JAN-1998 (TRENBLrel. 05, Last sequence update)  
 DT 01-NOV-1998 (TRENBLrel. 08, Last annotation update)  
 DE T02B11.6 PROTEIN.  
 GN T02B11.6.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;  
 OC Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX STRAIN-BRISTOL N2;  
 RX MEDLINE; 94150718.

RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,  
 RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,  
 RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,  
 RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,  
 RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,  
 RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,  
 RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,  
 RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULTON J.,  
 RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,  
 RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.,  
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
 RT elegans";  
 RL Nature 368:32-38(1994).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX STRAIN-BRISTOL N2;  
 RX GOELA D.;  
 RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.

RA WATERSTON R.;  
 RA Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF022979; AAB69904.1; -  
 DR PFAM; PF00083; sugar.tr. 1.  
 SQ SEQUENCE 466 AA; 51764 MW; 557F8291 CRC32;

Query Match 75.0%; Score 54; DB 5; Length 466;  
 Best Local Similarity 66.7%; Pred. No. 7.92e+00;  
 Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 302 LPFKFAAF 310  
 QY... 1 FPFKYAAA 9  
 :|||:||||

RESULT 7  
 ID Q9XWE8 PRELIMINARY; PRT; 470 AA.  
 AC Q9XWE8;  
 DT 01-NOV-1999 (TRENBLrel. 12, Created)  
 DT 01-NOV-1999 (TRENBLrel. 12, Last sequence update)  
 DT 01-NOV-1999 (TRENBLrel. 12, Last annotation update)  
 DE Y683B.5 PROTEIN.  
 GN Y683B.5.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;  
 OC Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX WHITE S.;  
 RX MEDLINE; 94150718.

RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
 RT elegans";  
 RL submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,  
 RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,  
 RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,  
 RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,  
 RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,  
 RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,  
 RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,  
 RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULTON J.,  
 RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,  
 RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.,  
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
 RT elegans";  
 RL Nature 368:32-38(1994).  
 DR EMBL; AL032655; CAA21725.1; -  
 SQ SEQUENCE 470 AA; 53242 MW; E8A2C32B CRC32;

Query Match 75.0%; Score 54; DB 5; Length 470;  
 Best Local Similarity 62.5%; Pred. No. 7.92e+00;  
 Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 434 FPFYSSA 441  
 QY 1 FPFKYAAA 8  
 |||||: 1

RESULT 8  
 ID Q46149 PRELIMINARY; PRT; 2178 AA.  
 AC Q46149; Q46147; Q46148;  
 DT 01-NOV-1996 (TRENBLrel. 01, Created)  
 DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)  
 DT 01-NOV-1999 (TRENBLrel. 12, Last annotation update)  
 DE ALPHA-TOXIN.  
 OS Clostridium novyi.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;  
 OC Clostridium.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX STRAIN-ATCC19402;  
 RX MEDLINE; 95342160.

RA HOFMANN F., HERRMANN A., HABERMANN E., VON EICHEL-STREIBER C.;  
 RT "Sequencing and analysis of the gene encoding the alpha-toxin of  
 RT Clostridium novyi proves its homology to toxins A and B of Clostridium  
 RT difficile";  
 RL Mol. Gen. Genet. 247:670-679(1995).  
 RN [2]  
 RP SEQUENCE OF 1204-2178 FROM N.A.  
 RC STRAIN-ATCC19402;  
 RA HOFMANN F., HABERMANN E., VON EICHEL-STREIBER C.;  
 RL Submitted (JUL-1993) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; Z48636; CAA88565.1; -  
 DR EMBL; Z23280; CAA80818.1; -



```

DR PFAM: PF00583; Acetyltransf; 1.
SQ SEQUENCE 173 AA; 20426 MW; 131A50A2 CRC32;

Query Match 76.4%; Score 55; DB 1; Length 173;
Best Local Similarity 55.6%; Pred. No. 5.18e+00;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

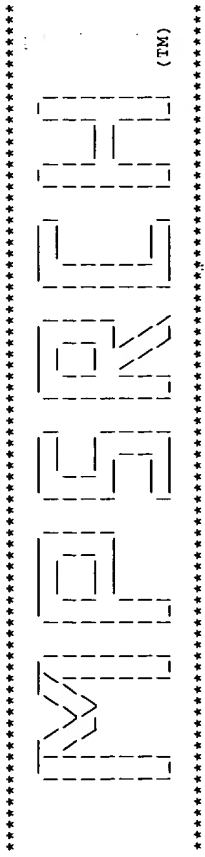
Db 32 PFFRYPLVF 40
   |||::|
QY 1 PPFKYAAAF 9

RESULT 4
ID P73738 PRELIMINARY; PRT; 469 AA.
AC P73738;
DT 01-FEB-1997 (TREMBlrel. 02, Created)
DT 01-FEB-1997 (TREMBlrel. 02, Last sequence update)
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
DE HYPOTHETICAL 52.5 KD PROTEIN.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-PCC6803;
RA TABATA S.;
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-PCC6803;
RX MEDLINE; 97081201.
RA KANEKO T., SATO S., KOTANI H., TANAKA A., ASAMIZU E., NAKAMURA Y.,
RA MIYAJIMA N., HIROSAWA M., SUGIURA M., SASAMOTO S., KIMURA T.,
RA HOSOUCHI T., MATSUNO A., MURAKI A., NAKAZAKI N., NARUO K., OKUMURA S.,
RA SHIMPO S., TAKEUCHI C., WADA T., WATANABE A., YAMADA M., YASUDA M.,
RA TABATA S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions.";
RL DNA Res. 3:109-136(1996).
DR EMBL; D90909; BAAL7786.1; -
DR PFAM; PF00355; Rieske; 1.
KW Hypothetical protein.
SQ SEQUENCE 469 AA; 52544 MW; 0887A65E CRC32;

Query Match 76.4%; Score 55; DB 2; Length 469;
Best Local Similarity 55.6%; Pred. No. 5.18e+00;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 282 PFFKPSKF 290
   |||::|
QY 1 PPFKYAAAF 9

RESULT 5
ID Q21145 PRELIMINARY; PRT; 800 AA.
AC Q21145;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-JAN-1999 (TREMBlrel. 09, Last annotation update)
DE K02E11.1 PROTEIN.
GN K02E11.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
OC Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN [1]
RP SEQUENCE FROM N.A.
RA MCMURRAY A.;
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 94150718.
RA WILSON R., AINSOUGH R., ANDERSON K., BAYNES C., BERKS M., COULSON A.,
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
```



Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Fri Apr 14 23:02:07 2000; MasPar time 14.04 Seconds  
44.432 Million cell updates/sec  
Tabular output not generated.  
Title: >US-08-452-843-1  
Description: (1-9) from US08452843.pep  
Perfect Score: 72  
Sequence: 1 FPFKYAAAF 9

Scoring table: PAM 150  
Gap 15  
Searched: 225878 seqs, 69334122 residues  
Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: spiremb12  
1:sp\_archaea 2:sp\_bacteria 3:sp\_fungi 4:sp\_human  
5:sp\_invertebrate 6:sp\_mammal 7:sp\_mhc 8:sp\_organelle  
9:sp\_phase 10:sp\_plant 11:sp\_rodent 12:sp\_unclassified  
13:sp\_vertebrate 14:sp\_virus

Statistics: Mean 23.533; Variance 33.793; scale 0.696  
Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES				Pred. No.	
Result No.	Score	Query Match	Description	ID	Pred. No.
1	60	83.3	H12C20.4 PROTEIN.	H12C20.4	5.78e-01
2	58	80.6	HYPOTHETICAL 58.7 KD P	Q9XXM7	1.41e+00
3	55	76.4	CONSERVED PROTEIN.	Q9Y256	5.18e+00
4	55	76.4	HYPOTHETICAL 52.5 KD P	P73738	5.18e+00
5	55	76.4	K02E11.1 PROTEIN.	Q21145	5.18e+00
6	54	75.0	T02B11.6 PROTEIN.	Q16976	7.92e+00
7	54	75.0	Y6B3B.5 PROTEIN.	Q9XW88	7.92e+00
8	54	75.0	ALPHA-TOXIN.	Q46149	7.92e+00
9	54	75.0	PKD1 PROTEIN.	Q42181	7.92e+00
10	53	73.6	HYPOTHETICAL 23.3 KD P	Q86755	1.21e+01
11	52	72.2	ZC518.3 PROTEIN.	Q23370	1.83e+01
12	52	72.2	KIAA0443.	Q43168	1.83e+01
13	51	70.8	RESTRICTION ENDONUCLEASE	Q23322	2.75e+01
14	51	70.8	HYPOTHETICAL 24.9 KD P	Q68567	2.75e+01
15	51	70.8	ZINC FINGER PROTEIN.	Q26053	2.75e+01
16	51	70.8	ZINC FINGER PROTEIN.	Q43434	2.75e+01
17	51	70.8	FIM PROTEIN.	Q60898	2.75e+01
18	50	69.4	HYPOTHETICAL 10.1 KD P	Q05234	4.13e+01
19	50	69.4	DNA-BINDING PROTEIN.	Q13863	4.13e+01
20	50	69.4	P190-C-MET (FRAGMENT).	Q63964	4.13e+01

21	50	69.4	252	2	Q9X9M5	PUTATIVE REPLICATION P	4.13e+01
22	50	69.4	321	5	O18046	T06C12.11 PROTEIN.	4.13e+01
23	50	69.4	329	4	Q9Y256	FARNESYLATED-PROTEINS	4.13e+01
24	50	69.4	440	5	O17358	C05E4.6 PROTEIN.	4.13e+01
25	50	69.4	453	2	O67436	PERIPLASMIC SERINE PRO	4.13e+01
26	50	69.4	490	2	O25863	NADH-UBIQUINONE OXIDOR	4.13e+01
27	50	69.4	492	2	Q92JV4	NADH OXIDOREDUCTASE I.	4.13e+01
28	50	69.4	511	2	O06460	BETA SUBUNIT OF NITRAT	4.13e+01
29	50	69.4	689	5	Q9Y114	PUTATIVE ZINC METALLOP	4.13e+01
30	50	69.4	1382	11	P97579	HEPATOCYTE GROWTH FACT	4.13e+01
31	50	69.4	1382	11	P97523	HGF RECEPTOR PRECURSOR	4.13e+01
32	49	68.1	139	10	O23746	POLLEN ALLERGEN, BETV1	6.15e+01
33	49	68.1	150	10	Q96370	POLLEN ALLERGEN BET V	6.15e+01
34	49	68.1	160	10	Q96366	POLLEN ALLERGEN BET V	6.15e+01
35	49	68.1	160	10	Q92S39	POLLEN ALLERGEN BETV1.	6.15e+01
36	49	68.1	160	10	O24642	POLLEN ALLERGEN BETV1.	6.15e+01
37	49	68.1	160	10	Q96365	POLLEN ALLERGEN BET V	6.15e+01
38	49	68.1	209	5	O01769	COSMID ZC581.	6.15e+01
39	49	68.1	291	1	O28106	4-HYDROXYBENZONATE OCTA	6.15e+01
40	49	68.1	292	13	Q92038	ACYL-COA DESATURASE (E	6.15e+01
41	49	68.1	481	5	Q96363	GRAM NEGATIVE BINDING	6.15e+01
42	49	68.1	607	2	O56974	KIM5.	6.15e+01
43	49	68.1	682	5	O44553	K06A5.1 PROTEIN.	6.15e+01
44	49	68.1	695	5	Q96441	XANTHINE DEHYDROGENASE	6.15e+01
45	49	68.1	822	1	O30286	MOLYBDOPTERIN OXIDORE	6.15e+01

ALIGNMENTS

RESULT	1	PRELIMINARY;	PRT;	286 AA.
ID	Q9XXM7			
AC	Q9XXM7			
DT	01-NOV-1999 (TREMELREL. 12, Created)			
DT	01-NOV-1999 (TREMELREL. 12, Last sequence update)			
DT	01-NOV-1999 (TREMELREL. 12, Last annotation update)			
DE	H12C20.4 PROTEIN.			
GN	H12C20.4.			
OS	Caenorhabditis elegans.			
OC	Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;			
OC	Rhabditina; Rhabditidae; Rhabditidae; Pelodidae; Pelodidae; Pelodidae;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE; 94150718.			
RA	WHITE S.;			
RT	"2.2 Mb of contiguous nucleotide sequence from chromosome III of C.			
RT	elegans";			
RL	Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE; 94150718.			
RA	BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,			
RA	CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,			
RA	GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,			
RA	JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,			
RA	LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,			
RA	PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,			
RA	SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,			
RA	THIERY-NIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,			
RA	WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;			
RT	"2.2 Mb of contiguous nucleotide sequence from chromosome III of C.			
RT	elegans";			
RL	Nature 368:32-38(1994).			
DR	EMBL; AL022272; CAA18352.1;			
SQ	SEQUENCE 286 AA; 33101 MW; 4B466B5E CRC32;			

Query Match 83.3%; Score 60; DB 5; Length 286;  
Best Local Similarity 66.7%; Pred. No. 5.78e-01;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
Db 87 FPFKYAAAF 95  
QY 1 FPFKYAAAF 9

THIS PAGE BLANK (USPTO)

FT VARIANT 1230 1230 Y -> C (IN HPRC; GERMLINE MUTATION).  
 FT FTID=VAR\_006292.  
 FT VARIANT 1230 1230 Y -> H (IN HPRC; SOMATIC MUTATION).  
 FT FTID=VAR\_006293.  
 FT VARIANT 1250 1250 M -> T (IN HPRC; SOMATIC MUTATION).  
 FT FTID=VAR\_006294.  
 FT CONFLICT 755 755 S -> STWKEPLNIVSFLFCFAS (IN REF. 2).  
 FT CONFLICT 1191 1191 G -> A (IN REF. 2).  
 SQ SEQUENCE 1390 AA: 155526 MW: 650992C2 CRC32;

Query Match 69.4%; Score 50; DB 1; Length 1390;  
 Best Local Similarity 55.6%; Pred. No. 1.47e+01;  
 Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Db 365 FPIKYVND 373  
 QY 1 FPFKYARAF 9

Search completed: Fri Apr 14 23:01:48 2000  
 Job time : 42 secs.

GN MET.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Primates; Catarrhini; Homiidae; Homo.  
RN [1]  
RA SEQUENCE FROM N.A.  
RP GIORDANO S.;  
RL Submitted (NOV-1990) to the EMBL/GenBank/DBJ databases.  
RX [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 87317655.  
RA PARK M., DEAN M., KAUL K., BRAUN M.J., GONDA M.A., VANDE WOUDE G.;  
RT "Sequence of MET protooncogene cDNA has features characteristic of  
the tyrosine kinase family of growth-factor receptors.";  
RN [3]  
RL proc. Natl. Acad. Sci. U.S.A. 84:6379-6383(1987).  
RP SEQUENCE OF 1010-1390 FROM N.A.  
RX MEDLINE; 88143699.  
RA CHAN A.M.L., KING H.W.S., TEMPEST P.R., DEAKIN E.A., COOPER C.S.,  
RA BROOKES P.;  
RT "Primary structure of the met protein tyrosine kinase domain.";  
RN [4]  
RL Oncogene 1:229-233(1987).  
RP SEQUENCE OF 1206-1264 FROM N.A.  
RX MEDLINE; 94067791.  
RA LEE S.T., STRUNK K.M., SPRITZ R.A.;  
RT "A survey of protein tyrosine kinase mRNAs expressed in normal human  
melanocytes.";  
RN [5]  
RL Oncogene 8:3403-3410(1993).  
RP SEQUENCE OF 1267-1390 FROM N.A.  
RX MEDLINE; 86063462.  
RA DEAN M., PARK M., LE BEAU M.M., ROBINS T.S., DIAZ M.O., ROWLEY J.D.,  
RA BLAIR D.G., VANDE WOUDE G.F.;  
RT "The human met oncogene is related to the tyrosine kinase oncogenes.";  
RN [6]  
RL Nature 318:385-388(1985).  
RP SEQUENCE OF 1-754 FROM N.A.  
RX PAULEY A., ANDREWS S.;  
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.  
RN [7]  
RP FUNCTION; 91118019.  
RA BOTTARO D.P., RUBIN J.S., FALETO D.L., CHAN A.M.-L., KMIECIK T.E.,  
RA VANDE WOUDE G.F., AARONSON S.A.;  
RT "Identification of the hepatocyte growth factor receptor as the c-met  
proto-oncogene product.";  
RN [8]  
RL Science 251:802-804(1991).  
RP PHOSPHORYLATION AT TYR-1235.  
RX MEDLINE; 92011756.  
RA FERRACINI R., LONGATI P., NALDINI L., VIGNA E., COMOGGIO P.M.;  
RT "Identification of the major autophosphorylation site of the  
Met/hepatocyte growth factor receptor tyrosine kinase.";  
RN [9]  
RL J. Biol. Chem. 266:19558-19564(1991).  
RP VARIANTS HPRC, AND VARIANT VAL-320.  
RX MEDLINE; 97285124.  
RA SCHMIDT L., DUH F.-M., CHEN F., KISHIDA T., GLENN G., CHOYE P.,  
RA SCHREER S.W., ZHUANG Z., LUBENSKY I., DEAN M., ALLIKMETS R.,  
RA CHIDAMARAM A., BERGERHEIM U.R., FELTIS J.T., CASADEVALL C.,  
RA ZAMARRON A., BERNUES M., RICHARD S., LIPS C.J.M., WALPHER M.M.,  
RA TSUI L.-C., GEIL L., ORCUTT M.L., STACKHOUSE T., LIPAN J., SLIFE L.,  
RA BRAUCH H., DECKER J., NIEHANS G., HUGHSON M.D., MOCH H., STORKEL S.,  
RA LERNAN M.I., LINHAN W.M., ZBAR B.;  
RT "Germline and somatic mutations in the tyrosine kinase domain of the  
MET proto-oncogene in papillary renal carcinomas.";  
RN [10]  
RL Nat. Genet. 16:68-73(1997).  
CC -1- FUNCTION: RECEPTOR FOR HEPATOCYTE GROWTH FACTOR. HAS A TYROSINE-  
PROTEIN KINASE ACTIVITY.  
CC -1- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE - ADP +  
PROTEIN TYROSINE PHOSPHATE.  
CC -1- SUBUNIT: HETERODIMER FORMED OF AN ALPHA CHAIN (50 KD) AND A BETA  
CHAIN (145 KD) WHICH ARE DISULFIDE LINKED.  
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.  
CC -1- DISEASE: ACTIVATION OF MET AFTER REARRANGEMENT WITH THE TPR  
GENE PRODUCES AN ONCOGENIC PROTEIN.  
CC -1- DISEASE: DEFECTS IN MET ARE THE CAUSE OF HEREDITARY PAPILLARY  
RENAL CARCINOMA (HPRC). HPRC IS A FORM OF INHERITED KIDNEY CANCER  
CHARACTERIZED BY A PREDISPOSITION TO DEVELOP MULTIPLE, BILATERAL  
PAPILLARY RENAL TUMORS. THE PATTERN OF INHERITANCE IS CONSISTENT  
WITH AUTOSOMAL DOMINANT TRANSMISSION WITH REDUCED PENETRANCE.  
CC -1- SIMILARITY: BELONGS TO THE MET TYROSINE KINASE FAMILY OF RECEPTOR.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC EMBL; M35074; AAA59590.1; -  
DR EMBL; X54559; CAB56793.1; -  
DR EMBL; J02958; AAA59591.1; -  
DR EMBL; AC002080; AAB54047.1; -  
DR PIR; A40175; TVHUME.  
DR HSP; P11362; IFGI.  
DR MIN; 164860; -  
DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
DR PROSITE; PS00109; PROTEIN\_KINASE\_TYR; 1.  
DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
DR PFAM; PF00069; Pkinase; 1.  
DR PFAM; PF01403; Sema; 1.  
DR PFAM; PF01437; Plexin-repeat; 1.  
DR Transferrase; Tyrosine-protein kinase; Proto-oncogene; ATP-binding;  
KW Receptor; Transmembrane; Glycoprotein; Phosphorylation; Signal;  
KW Chromosomal translocation; Disease mutation; Polymorphism.  
FT SIGNAL 1 24  
FT CHAIN 25 1390 HEPATOCYTE GROWTH FACTOR RECEPTOR.  
FT DOMAIN 25 932 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 933 955 POTENTIAL.  
FT DOMAIN 956 1390 CYTOPLASMIC (POTENTIAL).  
FT DOMAIN 1078 1345 PROTEIN KINASE.  
FT NP\_BIND 1084 1092 ATP (BY SIMILARITY).  
FT BINDING 1110 1110 ATP (BY SIMILARITY).  
FT ACT\_SITE 1204 1204 CLEAVAGE (POTENTIAL).  
FT SITE 307 308 BREAKPOINT FOR TRANSLOCATION TO FORM  
FT SITE 1009 1010 TPR-MET ONCOGENE.  
FT MOD\_RES 1235 1235 PHOSPHORYLATION (AUTO-).  
FT CARBOHYD 45 45 POTENTIAL.  
FT CARBOHYD 106 106 POTENTIAL.  
FT CARBOHYD 149 149 POTENTIAL.  
FT CARBOHYD 202 202 POTENTIAL.  
FT CARBOHYD 399 399 POTENTIAL.  
FT CARBOHYD 405 405 POTENTIAL.  
FT CARBOHYD 607 607 POTENTIAL.  
FT CARBOHYD 635 635 POTENTIAL.  
FT CARBOHYD 785 785 POTENTIAL.  
FT CARBOHYD 879 879 POTENTIAL.  
FT CARBOHYD 930 930 POTENTIAL.  
FT VARIANT 320 320 A -> V.  
FT VARIANT 1131 1131 /FTID=VAR\_006285.  
FT VARIANT 1188 1188 M -> T (IN HPRC; GERMLINE MUTATION).  
FT VARIANT 1195 1195 V -> L (IN HPRC; GERMLINE MUTATION).  
FT VARIANT 1220 1220 /FTID=VAR\_006286.  
FT VARIANT 1228 1228 L -> V (IN HPRC; SOMATIC MUTATION).  
FT VARIANT 1228 1228 /FTID=VAR\_006287.  
FT VARIANT 1228 1228 L -> V (IN HPRC; SOMATIC MUTATION).  
FT VARIANT 1228 1228 /FTID=VAR\_006288.  
FT VARIANT 1228 1228 V -> I (IN HPRC; GERMLINE MUTATION).  
FT VARIANT 1228 1228 D -> N (IN HPRC; GERMLINE MUTATION).  
FT VARIANT 1228 1228 /FTID=VAR\_006289.  
FT VARIANT 1228 1228 D -> H (IN HPRC; SOMATIC MUTATION).  
FT VARIANT 1228 1228 /FTID=VAR\_006290.  
FT VARIANT 1228 1228 /FTID=VAR\_006291.

CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
 CC EMBL; X64736; CAA46002.1; -;  
 CC EMBL; M97450; AAA29020.1; -;  
 CC DR FLYBASE; FBgn0010194; Wnt5.  
 CC DR PROSITE; PS00246; WNT1; 1.  
 CC DR PFAM; PF00110; wnt; 2.  
 CC KW Developmental protein; Glycoprotein; Signal.  
 CC FT SIGNAL 1 28  
 CC FT CHAIN 29 1010  
 CC FT DOMAIN 280 288  
 CC FT POLY-SER.  
 CC FT DOMAIN 461 476  
 CC FT CARBOHYD 60 60  
 CC FT POTENTIAL.  
 CC FT CARBOHYD 66 66  
 CC FT POTENTIAL.  
 CC FT CARBOHYD 115 115  
 CC FT POTENTIAL.  
 CC FT CARBOHYD 219 219  
 CC FT POTENTIAL.  
 CC FT CARBOHYD 310 310  
 CC FT POTENTIAL.  
 CC FT CARBOHYD 344 344  
 CC FT POTENTIAL.  
 CC FT CARBOHYD 425 425  
 CC FT POTENTIAL.  
 CC FT CARBOHYD 490 490  
 CC FT POTENTIAL.  
 CC FT CARBOHYD 491 491  
 CC FT POTENTIAL.  
 CC FT CARBOHYD 534 534  
 CC FT POTENTIAL.  
 CC FT CARBOHYD 599 599  
 CC FT POTENTIAL.  
 CC FT CARBOHYD 730 730  
 CC FT POTENTIAL.  
 CC FT CARBOHYD 958 958  
 CC FT POTENTIAL.  
 CC FT CONFLICT 281 283  
 CC FT MISSING (IN REF. 2).  
 CC FT CONFLICT 320 320  
 CC FT E -> D (IN REF. 2).  
 CC FT CONFLICT 474 476  
 CC FT MISSING (IN REF. 2).  
 CC SQ SEQUENCE 1010 AA; 112875 MW; 507BD98C CRC32;

Query Match 69.4%; Score 50; DB 1; Length 1010;  
 Best Local Similarity 44.4%; Pred. No. 1.47e+01;  
 Matches 4; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 673 FAYKFAATDF 681  
 |::|::|  
 QY 1 FPFKYAAAF 9

RESULT 14  
 ID MET\_MOUSE STANDARD; PRT; 1379 AA.  
 AC P16056; Q62125;  
 DT 01-APR-1990 (Rel. 14, Created)  
 DT 01-APR-1990 (Rel. 14, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE HEPATOCYTE GROWTH FACTOR RECEPTOR PRECURSOR (MET PROTO-ONCOGENE  
 DE TYROSINE KINASE) (EC 2.7.1.112) (HGF-SF RECEPTOR).  
 GN MET.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RP MEDLINE; 88262253.  
 RA CHAN A.M.L., KING H.W.S., DEAKIN E.A., TEMPEST P.R., HILKENS J.,  
 RA KROESEN V., EDWARDS D.R., WILLS A.J., BROOKES P., COOPER C.S.;  
 RT "Characterization of the mouse met proto-oncogene";  
 RL Oncogene 2:593-599(1988).  
 RN [2]  
 RP SEQUENCE OF 1199-1270 FROM N.A.  
 RP MEDLINE; 90152381.  
 RA WILKS A.F., KURBAN R.R., HOVENS C.M., RALPH S.J.;  
 RT "The application of the polymerase chain reaction to cloning members  
 RT of the protein tyrosine kinase family";  
 RL Gene 85:67-74(1989).  
 CC -1- FUNCTION: RECEPTOR FOR HEPATOCYTE GROWTH FACTOR. HAS A TYROSINE-  
 CC PROTEIN KINASE ACTIVITY.  
 CC -1- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE - ADP +  
 CC PROTEIN TYROSINE PHOSPHATE.

CC -1- SUBUNIT: HETERODIMER FORMED OF AN ALPHA CHAIN (50 KD) AND A BETA  
 CC CHAIN (145 KD) WHICH ARE DISULFIDE LINKED.  
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.  
 CC -1- DISEASE: ACTIVATION OF MET AFTER REARRANGEMENT WITH THE TPR  
 CC (TRANSLOCATED PROMOTER) LOCUS OF CHROMOSOME 1 PRODUCES AN  
 CC ONCOGENIC PROTEIN.  
 CC -1- SIMILARITY: BELONGS TO THE MET TYROSINE KINASE FAMILY OF RECEPTOR.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
 CC EMBL; Y00671; CAA68680.1; -;  
 CC EMBL; M33424; AAA40015.1; -;  
 CC DR PIR; S01254; S01254.  
 CC DR HSSP; P11362; IFGI.  
 CC DR MGD; MGI:96969; MET.  
 CC DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
 CC DR PROSITE; PS00109; PROTEIN\_KINASE\_TYR; 1.  
 CC DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
 CC DR PFAM; PF00069; pkinase; 1.  
 CC DR PFAM; PF01403; Sema; 1.  
 CC DR PFAM; PF01437; Plexin\_repeat; 1.  
 CC KW Transferase; Tyrosine-protein kinase; proto-oncogene; ATP-binding;  
 CC Receptor; transmembrane; Glycoprotein; Phosphorylation; Signal.  
 CC FT SIGNAL 1 24  
 CC FT CHAIN 25 1379  
 CC FT DOMAIN 25 931  
 CC FT TRANSMEM 932 934  
 CC FT POTENTIAL.  
 CC FT DOMAIN 955 1379  
 CC FT CYTOPLASMIC (POTENTIAL).  
 CC FT DOMAIN 1076 1343  
 CC FT PROTEIN KINASE.  
 CC FT SITE 306 307  
 CC FT CLEAVAGE (POTENTIAL).  
 CC FT NP\_BIND 1082 1090  
 CC FT ATP (BY SIMILARITY).  
 CC FT BINDING 1108 1108  
 CC FT ATP (BY SIMILARITY).  
 CC FT ACT\_SITE 1202 1202  
 CC FT BY SIMILARITY.  
 CC FT MOD\_RES 1233 1233  
 CC FT PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
 CC FT CARBOHYD 45 45  
 CC FT POTENTIAL.  
 CC FT CARBOHYD 106 106  
 CC FT POTENTIAL.  
 CC FT CARBOHYD 201 201  
 CC FT POTENTIAL.  
 CC FT CARBOHYD 357 357  
 CC FT POTENTIAL.  
 CC FT CARBOHYD 398 398  
 CC FT POTENTIAL.  
 CC FT CARBOHYD 404 404  
 CC FT POTENTIAL.  
 CC FT CARBOHYD 606 606  
 CC FT POTENTIAL.  
 CC FT CARBOHYD 634 634  
 CC FT POTENTIAL.  
 CC FT CARBOHYD 784 784  
 CC FT POTENTIAL.  
 CC FT CARBOHYD 878 878  
 CC FT POTENTIAL.  
 CC FT CONFLICT 1199 1199  
 CC FT V -> I (IN REF. 2).  
 CC FT CONFLICT 1255 1255  
 CC FT T -> R (IN REF. 2).  
 CC FT CONFLICT 1261 1261  
 CC FT K -> T (IN REF. 2).  
 CC FT CONFLICT 1269 1269  
 CC FT VL -> IP (IN REF. 2).  
 CC SQ SEQUENCE 1379 AA; 153548 MW; EL597FLA CRC32;

Query Match 69.4%; Score 50; DB 1; Length 1379;  
 Best Local Similarity 55.6%; Pred. No. 1.47e+01;  
 Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Db 364 FPKYVNDVF 372  
 |::|::|  
 QY 1 FPFKYAAAF 9

RESULT 15  
 ID MET\_HUMAN STANDARD; PRT; 1390 AA.  
 AC P08581;  
 DT 01-AUG-1988 (Rel. 08, Created)  
 DT 01-DEC-1992 (Rel. 24, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE HEPATOCYTE GROWTH FACTOR RECEPTOR PRECURSOR (MET PROTO-ONCOGENE  
 DE TYROSINE KINASE) (EC 2.7.1.112) (HGF-SF RECEPTOR).

-----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----

CC EMBL; M89778; AAA35211.1; -;  
 CC EMBL; X89633; CAA61776.1; -;  
 CC EMBL; Z75178; CAA99494.1; -;  
 CC EMBL; Z75179; CAA99496.1; -;  
 CC PIR; A42370; A42970.  
 CC SGD; L0002467; VPIL.  
 CC PFAM; PF01496; V-ATPase\_sub\_a; 1.  
 CC Hydrogen ion transport; Transmembrane; Glycoprotein.  
 CC DOMAIN 1 411  
 CC TRANSMEM 412 432  
 CC 1 (POTENTIAL).  
 CC DOMAIN 433 462  
 CC CYTOPLASMIC (POTENTIAL).  
 CC TRANSMEM 463 483  
 CC 2 (POTENTIAL).  
 CC DOMAIN 484 540  
 CC EXTRACELLULAR (POTENTIAL).  
 CC TRANSMEM 541 561  
 CC 3 (POTENTIAL).  
 CC DOMAIN 562 571  
 CC CYTOPLASMIC (POTENTIAL).  
 CC TRANSMEM 572 592  
 CC 4 (POTENTIAL).  
 CC DOMAIN 593 635  
 CC EXTRACELLULAR (POTENTIAL).  
 CC TRANSMEM 636 656  
 CC 5 (POTENTIAL).  
 CC DOMAIN 657 760  
 CC CYTOPLASMIC (POTENTIAL).  
 CC TRANSMEM 761 787  
 CC 6 (POTENTIAL).  
 CC DOMAIN 788 840  
 CC EXTRACELLULAR (POTENTIAL).  
 CC CARBOHYD 113 113  
 CC POTENTIAL.  
 CC CARBOHYD 280 280  
 CC POTENTIAL.  
 CC CARBOHYD 324 324  
 CC POTENTIAL.  
 CC SEQUENCE 840 AA; 95528 MW; A017D97B CRC32;

Query Match 69.4%; Score 50; DB 1; Length 840;  
 Best Local Similarity 55.6%; Pred. No. 1.47e+01;  
 Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 415 FPFYAAAF 423

QY 1 FPFYAAAF 9

RESULT 12

ID STV1 YEAST STANDARD; PRT; 890 AA.  
 AC P37296;  
 DT 01-OCT-1994 (Rel. 30, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 01-OCT-1996 (Rel. 34, Last annotation update)  
 DE VACUOLAR ATP SYNTHASE 101 KD SUBUNIT (EC 3.6.1.34) (V-ATPASE SUBUNIT AC115).  
 GN STV1 OR YMR054W OR YM9796.07.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;  
 OC Saccharomycetaceae; Saccharomycetes.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-X2180;  
 RX MEDLINE; 94245725.  
 RA MANOLSON M.F., WU B., PROTEAU D., TAILLON B.E., ROBERTS B.T.,  
 RA HOYT M.A., JONES E.W.;  
 RT "STV1 gene encodes functional homologue of 95-kDa yeast vacuolar  
 RT H(+)-ATPase subunit Vphlp.";  
 RL J. Biol. Chem. 269:14064-14074 (1994).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-S288C / AB972;  
 RA DEVLIN K., CHURCHER C.M., BARRELL B.G., RAJANDREAM M.A.;  
 RA Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: POTENTIAL ROLE IN DIFFERENTIAL TARGETING AND REGULATION  
 CC OF THE ENZYME FOR A SPECIFIC ORGANELLE.  
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. ENDOSOME.

CC -1- SIMILARITY: BELONGS TO THE V-ATPASE 116 KD SUBUNIT FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----

CC EMBL; U06465; AAA20596.1; -;  
 CC EMBL; Z49703; CAA89764.1; -;  
 CC PIR; A54081; A54081.  
 CC SGD; L0002139; SVIL.  
 CC PFAM; PF01496; V-ATPase\_sub\_a; 1.  
 CC Hydrogen ion transport; Transmembrane; Glycoprotein.  
 CC TRANSMEM 445 465  
 CC POTENTIAL.  
 CC TRANSMEM 467 487  
 CC POTENTIAL.  
 CC TRANSMEM 509 529  
 CC POTENTIAL.  
 CC TRANSMEM 585 605  
 CC POTENTIAL.  
 CC TRANSMEM 618 638  
 CC POTENTIAL.  
 CC TRANSMEM 681 701  
 CC POTENTIAL.  
 CC TRANSMEM 764 784  
 CC POTENTIAL.  
 CC TRANSMEM 796 816  
 CC POTENTIAL.  
 CC TRANSMEM 833 853  
 CC POTENTIAL.  
 CC CONFLICT 805 805  
 CC Q -> E (IN REF. 1).  
 CC SEQUENCE 890 AA; 101660 MW; F039D630 CRC32;

Query Match 69.4%; Score 50; DB 1; Length 890;  
 Best Local Similarity 55.6%; Pred. No. 1.47e+01;  
 Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 461 FPFYAAAF 469

QY 1 FPFYAAAF 9

RESULT 13

ID WNT5\_DROME STANDARD; PRT; 1010 AA.  
 AC P28466; Q01535;  
 DT 01-DEC-1992 (Rel. 24, Created)  
 DT 01-DEC-1992 (Rel. 24, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE WNT-5 PROTEIN PRECURSOR (DWT-5) (DWT-3).  
 GN WNT5 OR WNT-5 OR WNT-3.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CANTON-S;  
 RX MEDLINE; 93048811.  
 RA RUSSELL J., GENNITTSEN A., NUSSE R.;  
 RT "Isolation and expression of two novel Wnt/wingless gene homologues  
 RT in Drosophila.";  
 RL Development 115:475-482 (1992).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 93050786.  
 RA EISENBERG L.M., INGHAM P.W., BROWN A.M.C.;  
 RT "Cloning and characterization of a novel Drosophila Wnt gene, Dwt-5,  
 RT a putative downstream target of the homeobox gene distal-less.";  
 RL Dev. Biol. 154:73-83 (1992).  
 CC -1- FUNCTION: PROBABLE DEVELOPMENTAL PROTEIN. MAY BE A SIGNALING  
 CC MOLECULE WHICH AFFECTS THE DEVELOPMENT OF DISCRETE REGIONS OF  
 CC TISSUES. IS LIKELY TO SIGNAL OVER ONLY FEW CELL DIAMETERS. MAY  
 CC HAVE A ROLE IN THE CNS DEVELOPMENT.  
 CC -1- SUBCELLULAR LOCATION: SECRETED (PROBABLE).  
 CC -1- SIMILARITY: BELONGS TO THE WNT FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation.

GN PHBC.  
OS Alcaligenes eutrophus.  
OC Bacteria; Proteobacteria; beta subdivision; Burkholderia group;  
OC Ralstonia.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-H16;  
RA MEDLINE; 89359357.  
RX PEOPLES O.P., SINSKEY A.J.;  
RT "Poly-beta-hydroxybutyrate (PHB) biosynthesis in Alcaligenes  
eutrophus H16. Identification and characterization of the PHB  
polymerase gene (phbC).";  
RL J. Biol. Chem. 264:15298-15303(1989).  
RN [2]  
RP SEQUENCE OF 1-219 FROM N.A.  
RX MEDLINE; 91100279.  
RA SCHUBERT P., KRUGER N., STEINBUCH A.;  
RT "Molecular analysis of the Alcaligenes eutrophus  
poly(3-hydroxybutyrate) biosynthetic operon: Identification of the N  
terminus of poly(3-hydroxybutyrate) synthase and identification of  
the promoter";  
RL J. Bacteriol. 173:168-175(1991).  
CC -1- FUNCTION: POLYMERIZES D(-)-3-HYDROXYBUTYRYL-COA TO CREATE PHB  
WHICH SERVES AS AN INTRACELLULAR ENERGY RESERVE MATERIAL WHEN  
CELLS GROW UNDER CONDITIONS OF NUTRIENT LIMITATION.  
CC -1- PATHWAY: THIRD STEP IN POLY-BETA-HYDROXYBUTYRATE BIOSYNTHESIS.  
CC -1- SUBUNIT: MONOMER (PROBABLE).  
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.  
CC -1- SIMILARITY: TO OTHER PHA/PHB SYNTHASES.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC DE EMBL; J05003; AAA21975.1; -.  
DR EMBL; M64341; AAA21979.1; -.  
DR PIR; A34341; A34341.  
DR PIR; A34341; A34341.  
KW PHB biosynthesis; Transferase; Acyltransferase.  
FT ACT\_SITE 319 319 POTENTIAL.  
SQ SEQUENCE 589 AA: 64316 MW; FDCD5F11 CRC32;  
  
Query Match 69.4%; Score 50; DB 1; Length 589;  
Best Local Similarity 55.6%; Pred. No. 1.47e+01;  
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;  
  
Db 111 LPYRFAAAF 119  
QY 1 PPFKYAAAF 9  
  
RESULT 10  
ID Y366\_MYCPN STANDARD; PRT; 664 AA.  
AC P75234.  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE HYPOTHETICAL PROTEIN MG366 HOMOLOG.  
OS Mycoplasma pneumoniae.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;  
OC Mycoplasmataceae; Mycoplasma.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN-ATCC 29342 / M129;  
RX MEDLINE; 97105885.  
RA HIMMELREICH R., HILBERT H., PLAGENS H., PIRKL E., LI B.-C.,  
RA HERMANN R.;  
RT "Complete sequence analysis of the genome of the bacterium Mycoplasma  
pneumoniae.";

RL Nucleic Acids Res. 24:4420-4449(1996).  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC DE EMBL; AE000027; AAB95946.1; -.  
DR Hypothetical protein.  
SQ SEQUENCE 664 AA; 76769 MW; E4D94A05 CRC32;  
  
Query Match 69.4%; Score 50; DB 1; Length 664;  
Best Local Similarity 44.4%; Pred. No. 1.47e+01;  
Matches 4; Conservative 3; Mismatches 2; Indels 0; Gaps 0;  
  
Db 358 FAYKSEIF 366  
QY 1 PPFKYAAAF 9  
  
RESULT 11  
ID VPH1\_YEAST STANDARD; PRT; 840 AA.  
AC P32563;  
DT 01-OCT-1993 (Rel. 27, Created)  
DT 01-OCT-1993 (Rel. 27, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE VACUOLAR ATP SYNTHASE 95 KD SUBUNIT (EC 3.6.1.34) (VACUOLAR ATPASE 95  
KD SUBUNIT).  
GN VPH1 OR VOR270C.  
OS Saccharomycetes cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;  
OC Saccharomycetaceae; Saccharomycetes.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 92332542.  
RA MANOLSON M.F., PROTEAU D., PRESTON R.A., STENBIT A., ROBERTS B.T.,  
RA HOYT M.A., PREUSS D., MULHOLLAND J., BOTSTEIN D., JONES E.W.;  
RT "The VPH1 gene encodes a 95-kDa integral membrane polypeptide  
required for in vivo assembly and activity of the yeast vacuolar  
H(+)-ATPase.";  
RL J. Exp. Biol. 172:105-112(1992).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 93147685.  
RA MANOLSON M.F., PROTEAU D., JONES E.W.;  
RT "Evidence for a conserved 95-120 kDa subunit associated with and  
essential for activity of V-ATPases.";  
RL J. Exp. Biol. 172:105-112(1992).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-S288C;  
RX MEDLINE; 97051594.  
RA CHERET G., BERNARDI A., SOR F.J.;  
RT "DNA sequence analysis of the VPH1-SNF2 region on chromosome XV of  
Saccharomycetes cerevisiae.";  
RL Yeast 12:1059-1064(1996).  
RN [4]  
RP SEQUENCE FROM N.A.  
RC STRAIN-S288C / FY1679;  
RX MEDLINE; 97298311.  
RA POIREY R., JAUNIAUX J.C.;  
RT "Sequencing analysis of a 36.8 kb fragment of yeast chromosome XV  
reveals 26 open reading frames including SEC63, CDC31, SUG2, GCD1,  
RBL2, PNT1, PAC1 and VPH1.";  
RL Yeast 13:483-487(1997).  
CC -1- FUNCTION: REQUIRED FOR ASSEMBLY AND ACTIVITY OF THE VACUOLAR  
ATPASE. POTENTIAL ROLE IN DIFFERENTIAL TARGETING AND REGULATION OF  
THE ENZYME FOR A SPECIFIC ORGANELLE.  
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. VACUOLAR.  
CC -1- SIMILARITY: BELONGS TO THE V-ATPASE 116 KD SUBUNIT FAMILY.



Best Local Similarity 62.5%; Pred. No. 9.58e+00; Mismatches 2; Indels 0; Gaps 0;

Db 526 YPFKYAVS 533  
QY 1 FPFKYAAA 8

RESULT 7  
ID NMT\_CANAL STANDARD; PRT; 451 AA.  
AC P30418;  
DT 01-APR-1993 (Rel. 25, Created)  
DT 01-APR-1993 (Rel. 25, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE GLYCYPEPTIDE N-TETRADECANOYLTRANSFERASE (EC 2.3.1.97) (PEPTIDE  
DE N-MYRISTOYLTRANSFERASE) (MYRISTOYL-COA:PROTEIN N-MYRISTOYLTRANSFERASE)  
DE (NMT).  
GN NMT1  
OS Candida albicans (Yeast).  
OC Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;  
OC Candidaceae; Candida.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 92235090.  
RA WIEGAND R.C., CARR C., MINNERLY J.C., PAULEY A.M., CARRON C.P.,  
RA LANGNER C.A., TUCKER A.D., TUNNICLIFFE A., MISTRY A., MANCIA F.,  
RT "The Candida albicans myristoyl-CoA:protein N-myristoyltransferase  
RT gene. Isolation and expression in Saccharomyces cerevisiae and  
RT Escherichia coli.";  
RL J. Biol. Chem. 267:8591-8598(1992).  
RN [2]  
RP X-RAY CRYSTALLOGRAPHY (2.45 ANGSTROMS).  
RX MEDLINE: 98162557.  
RA WESTON S.A., CAMBLE R., COLLS J., ROSENBROCK G., TAYLOR I.,  
RA EGERTON M., TUCKER A.D., TUNNICLIFFE A., MISTRY A., MANCIA F.,  
RA DE LA FORTELLE E., IRWIN J., BRICOGNE G., PAUPTIT R.A.;  
RT "Crystal structure of the anti-fungal target N-myristoyl  
RT transferase.";  
RL Nat. Struct. Biol. 5:213-221(1998).  
RN [3]  
RP FUNCTION: ADDS MYRISTOYL GROUP TO N-TERMINAL GLYCINE RESIDUE  
OF CERTAIN CELLULAR AND VIRAL PROTEINS.  
CC -1- CATALYTIC ACTIVITY: TETRADECANOYL-COA + GLYCYL-PEPTIDE = COA +  
CC N-TETRADECANOYLGLYCYL-PEPTIDE.  
CC -1- SUBUNIT: MONOMER.  
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.  
CC -1- SIMILARITY: BELONGS TO THE NMT FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: M80544; AAA34351.1; -;  
DR PIR: A38099; A38099.  
DR PDB: 1NMT; 16-FEB-99.  
DR PROSITE: PS00975; NMT\_1; 1.  
DR PROSITE: PS00976; NMT\_2; 1.  
DR PFAM: PF01233; NMT; 1.  
KW Transferase; Acyltransferase; 3D-structure.  
SQ SEQUENCE 451 AA; 51877 MW; 52BD42D9 CRC32;

Query Match 69.4%; Score 50; DB 1; Length 451;  
Best Local Similarity 55.6%; Pred. No. 1.47e+01;  
Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Db 115 FRFKYSHEF 123  
QY 1 FPFKYAAA 9

RESULT 8  
ID YJCG\_ECOLI STANDARD; PRT; 549 AA.  
AC P32705;  
DT 01-OCT-1993 (Rel. 27, Created)  
DT 01-OCT-1993 (Rel. 27, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE HYPOTHETICAL 59.2 KD PROTEIN IN SOXR-ACS INTERGENIC REGION (F549).  
GN YJCG.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Escherichia.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-K12 / MG1655;  
RX MEDLINE: 94089392.  
RA BLATTNER F.R., BURLAND V.D., PLUNKETT G. III, SOFIA H.J.,  
RA DANIELS D.L.;  
RT "Analysis of the Escherichia coli genome. IV. DNA sequence of the  
RT region from 89.2 to 92.8 minutes.";  
RL Nucleic Acids Res. 21:5408-5417(1993).  
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).  
CC -1- SIMILARITY: BELONGS TO THE SODIUM: SODIUM SYMPORTER FAMILY (SSF).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: U00006; AAC43161.1; -;  
DR EMBL: AE000480; AAC77037.1; -;  
DR ECOGENE: EG11942; YJCG.  
DR PROSITE: PS00456; NA\_SOLUT\_SYMP\_1; 1.  
DR PROSITE: PS00457; NA\_SOLUT\_SYMP\_2; 1.  
DR PFAM: PF00474; SSF; 1.  
KW Hypothetical protein; Transport; Transmembrane; Sodium transport;  
KW Symport.  
FT TRANSMEM 4 24 POTENTIAL.  
FT TRANSMEM 33 53 POTENTIAL.  
FT TRANSMEM 77 97 POTENTIAL.  
FT TRANSMEM 103 123 POTENTIAL.  
FT TRANSMEM 148 168 POTENTIAL.  
FT TRANSMEM 183 203 POTENTIAL.  
FT TRANSMEM 206 226 POTENTIAL.  
FT TRANSMEM 262 282 POTENTIAL.  
FT TRANSMEM 303 323 POTENTIAL.  
FT TRANSMEM 355 375 POTENTIAL.  
FT TRANSMEM 404 424 POTENTIAL.  
FT TRANSMEM 428 448 POTENTIAL.  
FT TRANSMEM 464 484 POTENTIAL.  
FT TRANSMEM 493 513 POTENTIAL.  
SQ SEQUENCE 549 AA; 59197 MW; 2F07CF2F CRC32;

Query Match 69.4%; Score 50; DB 1; Length 549;  
Best Local Similarity 55.6%; Pred. No. 1.47e+01;  
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 494 FPEYEDALF 502  
QY 1 FPFKYAAA 9

RESULT 9  
ID PHBC\_ALCEU STANDARD; PRT; 589 AA.  
AC P23608;  
DT 01-NOV-1991 (Rel. 20, Created)  
DT 01-NOV-1991 (Rel. 20, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE POLY-BETA-HYDROXYBUTYRATE POLYMERASE (EC 2.3.1.-) (PHB POLYMERASE)  
DE (POLY(3-HYDROXYALKANOATE) POLYMERASE) (PHA-POLYMERASE) (PHA SYNTHASE)  
DE (POLYHYDROXYALKANOIC ACID SYNTHASE).

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).  
-----  
DR EMBL; Z35932; CAA85006.1; -  
DR PIR; S45923; S45923.  
KW Hypothetical protein; Transmembrane.  
FT TRANSMEM 35 55 POTENTIAL.  
FT TRANSMEM 92 112 POTENTIAL.  
SQ SEQUENCE 404 AA; 46444 MW; 821F8780 CRC32;  
  
Query Match 70.8%; Score 51; DB 1; Length 404;  
Best Local Similarity 44.4%; Pred. No. 9.58e+00;  
Matches 4; Conservative 3; Mismatches 2; Indels 0; Gaps 0;  
  
Db 10 PPFYGSDF 18  
||: ||: |  
QY 1 PPFYAAAF 9  
  
RESULT 5  
ID CSG.METFE STANDARD; PRT; 593 AA.  
AC P27373;  
DT 01-AUG-1992 (Rel. 23, Created)  
DT 01-AUG-1992 (Rel. 23, Last sequence update)  
DT 01-JUN-1994 (Rel. 23, Last annotation update)  
DE CELL SURFACE GLYCOPROTEIN PRECURSOR (S-LAYER PROTEIN).  
GN SLGA.  
OS Methanothermus fervidus.  
OC Archaea; Euryarchaeota; Methanobacteriales; Methanothermaceae;  
OC Methanothermus.  
RN [1]  
RP SEQUENCE FROM N.A., AND SEQUENCE OF 23-42.  
RC STRAIN-DSM 2088 / V24S;  
RX MEDLINE; 91293115.  
RA BROECKL G., BEHR M., FABRY S., HENSEL R., KAUEWITZ H., BIENDL E.,  
RA KOENIG H.;  
RT "Analysis and nucleotide sequence of the genes encoding the surface-  
RT layer glycoproteins of the hyperthermophilic methanogens  
RT Methanothermus fervidus and Methanothermus sociabilis.";  
RL Eur. J. Biochem. 199;147-152(1991).  
CC -!- FUNCTION: THE S-LAYER IS A PARACRYSTALLINE MONO-LAYERED ASSEMBLY  
CC OF PROTEINS WHICH COAT THE SURFACE OF BACTERIA.  
CC -!- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIA IS COVERED BY A  
CC S-LAYER WITH HEXAGONAL SYMMETRY.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL; X58297; CAA41230.1; -  
DR PIR; S16225; S16225.  
KW Glycoprotein; Cell wall; S-layer; Signal.  
FT SIGNAL 1 22  
FT CHAIN 23 593 CELL SURFACE GLYCOPROTEIN.  
FT CARBOHYD 29 29  
FT CARBOHYD 58 58 POTENTIAL.  
FT CARBOHYD 66 66 POTENTIAL.  
FT CARBOHYD 74 74 POTENTIAL.  
FT CARBOHYD 114 114 POTENTIAL.  
FT CARBOHYD 122 122 POTENTIAL.  
FT CARBOHYD 145 145 POTENTIAL.  
FT CARBOHYD 148 148 POTENTIAL.  
FT CARBOHYD 158 158 POTENTIAL.  
FT CARBOHYD 176 176 POTENTIAL.  
FT CARBOHYD 208 208 POTENTIAL.  
FT CARBOHYD 231 231 POTENTIAL.  
FT CARBOHYD 326 326 POTENTIAL.  
FT CARBOHYD 336 336 POTENTIAL.  
FT CARBOHYD 340 340 POTENTIAL.  
FT CARBOHYD 431 431 POTENTIAL.  
SQ SEQUENCE 431 AA; 46444 MW; 821F8780 CRC32;  
  
Query Match 70.8%; Score 51; DB 1; Length 404;  
Best Local Similarity 44.4%; Pred. No. 9.58e+00;  
Matches 4; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

FT CARBOHYD 471 471 POTENTIAL.  
FT CARBOHYD 500 500 POTENTIAL.  
FT CARBOHYD 516 516 POTENTIAL.  
SQ SEQUENCE 593 AA; 65481 MW; 5CFA9AA9 CRC32;  
  
Query Match 70.8%; Score 51; DB 1; Length 593;  
Best Local Similarity 62.5%; Pred. No. 9.58e+00;  
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
  
Db 526 YPFKYAVS 533  
:|||||:  
QY 1 PPFYAAAF 8  
  
RESULT 6  
ID CSG.METSC STANDARD; PRT; 593 AA.  
AC P27374;  
DT 01-AUG-1992 (Rel. 23, Created)  
DT 01-AUG-1992 (Rel. 23, Last sequence update)  
DT 01-JUN-1994 (Rel. 29, Last annotation update)  
DE CELL SURFACE GLYCOPROTEIN PRECURSOR (S-LAYER PROTEIN).  
GN SLGA.  
OS Methanothermus sociabilis.  
OC Archaea; Euryarchaeota; Methanobacteriales; Methanothermaceae;  
OC Methanothermus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-DSM 3496 / KF1-FL;  
RX MEDLINE; 91293115.  
RA BROECKL G., BEHR M., FABRY S., HENSEL R., KAUEWITZ H., BIENDL E.,  
RA KOENIG H.;  
RT "Analysis and nucleotide sequence of the genes encoding the surface-  
RT layer glycoproteins of the hyperthermophilic methanogens  
RT Methanothermus fervidus and Methanothermus sociabilis.";  
RL Eur. J. Biochem. 199;147-152(1991).  
CC -!- SUBUNIT: ASSEMBLY INTO MONO-LAYERED CRYSTALLINE ARRAYS.  
CC -!- SUBCELLULAR LOCATION: CELL WALL.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL; X58296; CAA41229.1; -  
DR PIR; S16375; S16375.  
KW Glycoprotein; Cell wall; S-layer; Signal.  
FT SIGNAL 1 22  
FT CHAIN 23 593 CELL SURFACE GLYCOPROTEIN.  
FT CARBOHYD 29 29  
FT CARBOHYD 58 58 POTENTIAL.  
FT CARBOHYD 66 66 POTENTIAL.  
FT CARBOHYD 74 74 POTENTIAL.  
FT CARBOHYD 114 114 POTENTIAL.  
FT CARBOHYD 122 122 POTENTIAL.  
FT CARBOHYD 145 145 POTENTIAL.  
FT CARBOHYD 148 148 POTENTIAL.  
FT CARBOHYD 158 158 POTENTIAL.  
FT CARBOHYD 176 176 POTENTIAL.  
FT CARBOHYD 208 208 POTENTIAL.  
FT CARBOHYD 231 231 POTENTIAL.  
FT CARBOHYD 326 326 POTENTIAL.  
FT CARBOHYD 336 336 POTENTIAL.  
FT CARBOHYD 340 340 POTENTIAL.  
FT CARBOHYD 431 431 POTENTIAL.  
FT CARBOHYD 471 471 POTENTIAL.  
FT CARBOHYD 500 500 POTENTIAL.  
FT CARBOHYD 516 516 POTENTIAL.  
SQ SEQUENCE 593 AA; 65503 MW; 52B1B8C8 CRC32;  
  
Query Match 70.8%; Score 51; DB 1; Length 593;

Query Match 72.2%; Score 52; DB 1; Length 503;  
Best Local Similarity 55.6%; Pred. No. 6.22e+00;  
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

DB 23 FAFNYVAGF 31  
|:|:| |:|  
QY 1 PPFKYAAAF 9

RESULT 2  
ID NMT\_AJCA STANDARD; PRT; 529 AA.  
AC P34763;  
DT 01-FEB-1994 (Rel. 28, Created)  
DT 01-FEB-1994 (Rel. 28, Last sequence update)  
DT 01-JUN-1994 (Rel. 29, Last annotation update)  
DE GLYCYLPEPTIDE N-TETRADECANOYLTRANSFERASE (EC 2.3.1.97) (PEPTIDE  
DE N-MYRISTOYLTRANSFERASE) (MYRISTOYL-COA:PROTEIN N-MYRISTOYLTRANSFERASE)  
DE (NMT).  
OS Ajellomyces capsulata (Histoplasma capsulatum).  
OC Eukaryota; Fungi; Ascomycota; Euascomycetes; Plectomycetes;  
OC Onygenales; Onygenaceae; Ajellomyces.  
[1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-G217B;  
RX MEDLINE; 94132075.  
RA LODGE J.K., JOHNSON R.L., WEINBERG R.A., GORDON J.I.;  
RT "Comparison of myristoyl-CoA:protein N-myristoyltransferases from  
RT three pathogenic fungi: Cryptococcus neoformans, Histoplasma  
RT capsulatum, and Candida albicans.";  
RL J. Biol. Chem. 269:2996-3009(1994).  
CC -|- FUNCTION: ADDS MYRISTOYL GROUP TO N-TERMINAL GLYCINE RESIDUE  
CC OF CERTAIN CELLULAR AND VIRAL PROTEINS.  
CC -|- CATALYTIC ACTIVITY: TETRADECANOYL-COA + GLYCYL-PEPTIDE - COA +  
CC N-TETRADECANOYLGLYCYL-PEPTIDE.  
CC -|- SUBCELLULAR LOCATION: CYTOPLASMIC.  
CC -|- SIMILARITY: BELONGS TO THE NMT FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DB EMBL; L25118; AAA17549.1; -;  
DR PROSITE; PS00975; NMT\_1; 1.  
DR PROSITE; PS00976; NMT\_2; 1.  
DR PFAM; PF01233; NMT; 1.  
KW Transferase; Acyltransferase.  
SQ SEQUENCE 529 AA; 59363 MW; 6B6ED646 CRC32;

Query Match 72.2%; Score 52; DB 1; Length 529;  
Best Local Similarity 55.6%; Pred. No. 6.22e+00;  
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

DB 192 FRFNYSPAF 200  
|:|:| |:|  
QY 1 PPFKYAAAF 9

RESULT 3  
ID ARYA\_MANSE STANDARD; PRT; 702 AA.  
AC P14296;  
DT 01-JAN-1990 (Rel. 13, Created)  
DT 01-JAN-1990 (Rel. 13, Last sequence update)  
DT 01-FEB-1994 (Rel. 28, Last annotation update)  
DE ARYLPHORIN ALPHA SUBUNIT PRECURSOR.  
OS Manduca sexta (Tobacco hawkmoth) (Tobacco hornworm).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;  
OC Sphingioidea; Sphingidae; Sphinginae; Manduca.  
[1]

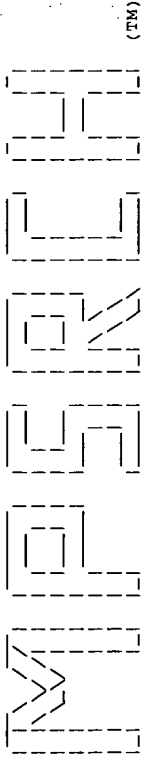
SEQUENCE FROM N.A.  
RC TISSUE-LARVAL FAT BODY;  
RX MEDLINE; 90037032.  
RA WILLOTT E., WANG X.-Y., WELLS M.A.;  
RT "cDNA and gene sequence of Manduca sexta arylphorin, an aromatic  
RT amino acid-rich larval serum protein. Homology to arthropod  
RT hemocyanins.";  
RL J. Biol. Chem. 264:19052-19059(1989).  
CC -|- FUNCTION: ARYLPHORIN IS A LARVAL STORAGE PROTEIN (LSP) WHICH MAY  
CC SERVE AS A STORAGE PROTEIN USED PRIMARILY AS A SOURCE OF AROMATIC  
CC AMINO ACIDS FOR PROTEIN SYNTHESIS DURING METAMORPHOSIS. IT IS A  
CC CONSTITUENT OF THE SCLEROTIZING SYSTEM OF THE CUTICLE, AND SERVES  
CC AS A CARRIER FOR ECDYSTEROID HORMONE.  
CC -|- SUBUNIT: ARYLPHORIN IS AN HEXAMER OF SUBUNITS ALPHA AND BETA.  
CC -|- TISSUE SPECIFICITY: FAT BODY.  
CC -|- SIMILARITY: TO ARYLPHORIN, TO B-MORI STORAGE PROTEINS 1 AND 2, AND TO  
CC ARTHROPOD HEMOCYANINS.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DB EMBL; M28394; AAA28302.1; -;  
DR EMBL; M28396; AAA29303.1; -;  
DR PIR; A34434; A34434.  
DR HSSP; P04253; 1LL1.  
DR PROSITE; PS00209; HEMOCYANIN\_1; 1.  
DR PROSITE; PS00210; HEMOCYANIN\_2; 1.  
DR PFAM; PF00372; hemocyanin; 1.  
KW Signal; Storage protein; Glycoprotein; Multigene family.  
FT SIGNAL 1 16  
FT CHAIN 17 702 ARYLPHORIN ALPHA SUBUNIT.  
FT CARBOHYD 75 75 POTENTIAL.  
FT CARBOHYD 214 214 POTENTIAL.  
SQ SEQUENCE 702 AA; 83866 MW; 5F4E87CD CRC32;

Query Match 72.2%; Score 52; DB 1; Length 702;  
Best Local Similarity 44.4%; Pred. No. 6.22e+00;  
Matches 4; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

DB 686 PFYKFNVPF 694  
|:|:| |:|  
QY 1 PPFKYAAAF 9

RESULT 4  
ID YBR3\_YEAST STANDARD; PRT; 404 AA.  
AC P38083;  
DT 01-OCT-1994 (Rel. 30, Created)  
DT 01-OCT-1994 (Rel. 30, Last sequence update)  
DT 01-NOV-1995 (Rel. 32, Last annotation update)  
DE HYPOTHETICAL 46.4 KD PROTEIN IN ORC2-TIP1 INTERGENIC REGION.  
GN YBR063C OR YBR0610.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;  
OC Saccharomycetaceae; Saccharomycetes.  
[1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-S288C;  
RA DOMDEY H., GASSENHUBER H., OBERMAIER B., PIRAVANDI E.;  
RL Submitted (AUG-1994) to the EMBL/GenBank/DBJ databases.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----

\*\*\*\*\*



\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MParch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Fri Apr 14 23:01:06 2000; MasPar time 4.22 Seconds  
Tabular output not generated. 63.743 Million cell updates/sec

Title: >US-08-452-843-1  
Description: (1-9) from US08452843.pep  
Perfect Score: 72  
Sequence: 1 FPFYAAAF 9

Scoring table: PAM 150  
Gap 15

Searched: 82229 seqs, 29864866 residues

Post-processing: Minimum Watch 0%  
Listing first 45 summaries

Database: swiss-prot38  
1:swissprot

Statistics: Mean 24.416; Variance 32.258; scale 0.757

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	52	72.2	503	1 Y226_MYCPN	HYPOTHETICAL PROTEIN M	6.22e+00
2	52	72.2	529	1 NMT_AJEC	GLYCYLPEPTIDE N-TETRAD	6.22e+00
3	52	72.2	702	1 ARYA_MANSE	ARYLPHORIN ALPHA SUBUN	6.22e+00
4	51	70.8	404	1 YBR3_YEAST	HYPOTHETICAL 46.4 KD P	9.58e+00
5	51	70.8	593	1 CSG_METFE	CELL SURFACE GLYCOPROT	9.58e+00
6	51	70.8	593	1 CSG_METSC	CELL SURFACE GLYCOPROT	9.58e+00
7	50	69.4	451	1 NMT_CANAL	GLYCYLPEPTIDE N-TETRAD	1.47e+01
8	50	69.4	549	1 XJCG_ECOLI	HYPOTHETICAL 59.2 KD P	1.47e+01
9	50	69.4	589	1 PHBC_ALCEU	POLY-BETA-HYDROXYBUTYR	1.47e+01
10	50	69.4	664	1 Y366_MYCPN	HYPOTHETICAL PROTEIN M	1.47e+01
11	50	69.4	840	1 VPH1_YEAST	VACUOLAR ATP SYNTHASE	1.47e+01
12	50	69.4	890	1 STV1_YEAST	VACUOLAR ATP SYNTHASE	1.47e+01
13	50	69.4	1010	1 WNT5_DROME	WNT-5 PROTEIN PRECURSOR	1.47e+01
14	50	69.4	1379	1 MET_MOUSE	HEPATOCYTE GROWTH FACT	1.47e+01
15	50	69.4	1390	1 MET_HUMAN	HEPATOCYTE GROWTH FACT	1.47e+01
16	49	68.1	131	1 MCRD_METFE	METHYL-COENZYME M REDU	2.23e+01
17	49	68.1	159	1 BVIF_BETVE	MAJOR POLLEN ALLERGEN	2.23e+01
18	49	68.1	159	1 BVIG_BETVE	MAJOR POLLEN ALLERGEN	2.23e+01
19	49	68.1	159	1 BV1J_BETVE	MAJOR POLLEN ALLERGEN	2.23e+01
20	49	68.1	159	1 BV1A_BETVE	MAJOR POLLEN ALLERGEN	2.23e+01
21	49	68.1	159	1 BV1D_BETVE	MAJOR POLLEN ALLERGEN	2.23e+01
22	49	68.1	159	1 BV1L_BETVE	MAJOR POLLEN ALLERGEN	2.23e+01
23	49	68.1	308	1 IFRH_SOLTU	ISOFLAVONE REDUCTASE H	2.23e+01

24	49	68.1	355	1 ACOL_MOUSE	ACYL-COA DESATURASE 1	2.23e+01
25	49	68.1	427	1 Y96E_MYCPN	HYPOTHETICAL PROTEIN M	2.23e+01
26	49	68.1	450	1 NMT_CAEEL	PROBABLE GLYCYLPEPTIDE	2.23e+01
27	49	68.1	545	1 VNGS_JCDNV	NONCAPSID PROTEIN NS-1	2.23e+01
28	49	68.1	607	1 YSCC_YEREN	YOP PROTEINS TRANSLOCA	2.23e+01
29	49	68.1	701	1 HRPH_PSEY	HYPERSENSITIVITY RESPO	2.23e+01
30	49	68.1	735	1 YDD8_SCHPO	HYPOTHETICAL 83.3 KD P	2.23e+01
31	49	68.1	1342	1 XDH_DROPS	XANTHINE DEHYDROGENASE	2.23e+01
32	49	68.1	1344	1 XDH_DROSU	XANTHINE DEHYDROGENASE	2.23e+01
33	48	66.7	585	1 YH70_SYNY3	HYPOTHETICAL 67.1 KD P	3.36e+01
34	48	66.7	655	1 HGFA_HUMAN	HEPATOCYTE GROWTH FACT	3.36e+01
35	48	66.7	683	1 APCE_SYNP6	PHYCOBILISOME LINKER P	3.36e+01
36	48	66.7	1131	1 PHA_SOYBN	PHYTOCHROME A.	3.36e+01
37	48	66.7	1353	1 XDH_CALVI	XANTHINE DEHYDROGENASE	3.36e+01
38	48	66.7	1354	1 PUR4_DROME	PHOSPHORIBOSYLFORMYLGL	5.08e+01
39	47	65.3	308	1 Y222_MYCPN	HYPOTHETICAL PROTEIN M	5.08e+01
40	47	65.3	365	1 GAL7_YEAST	GALACTOSE-1-PHOSPHATE	5.08e+01
41	47	65.3	751	1 PANG_DROME	PROTEIN PANGOLIN.	5.08e+01
42	47	65.3	764	1 YJJO_YEAST	HYPOTHETICAL 87.2 KD P	5.08e+01
43	47	65.3	1125	1 PHVA_POPTM	PHYTOCHROME A.	5.08e+01
44	47	65.3	1129	1 PHVB_SOLTU	PHYTOCHROME B.	5.08e+01
45	47	65.3	1171	1 PHVB_ORYSA	PHYTOCHROME B.	5.08e+01

ALIGNMENTS

RESULT	ID	Y226_MYCPN	STANDARD;	PRT;	503 AA.
AC	P75462:				
DT	01-NOV-1997 (Rel. 35, Created)				
DT	01-NOV-1997 (Rel. 35, Last sequence update)				
DE	HYPOTHETICAL PROTEIN MG226 HOMOLOG.				
OS	Mycoplasma pneumoniae.				
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;				
CC	Mycoplasmataceae; Mycoplasma.				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN-ATCC 29342 / M129;				
RX	MEDLINE: 97105885				
RA	HIMMELREICH R., HILBERT H., FLAGENS H., PIRKL E., LI B.-C.,				
RA	HERMANN R.;				
RT	"Complete sequence analysis of the genome of the bacterium Mycoplasma pneumoniae."				
RL	Nucleic Acids Res. 24:4420-4449(1996).				
CC	- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).				
CC	- SIMILARITY: TO M.GENITALIUM MG225.				

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----

DR	AE000051; AAB96165.1;				
DR	PFAM: PF00324; aa_permeases; 1.				
KW	Hypothetical protein; Transmembrane.				
FT	TRANSMEM 20 40				POTENTIAL.
FT	TRANSMEM 43 63				POTENTIAL.
FT	TRANSMEM 106 126				POTENTIAL.
FT	TRANSMEM 138 158				POTENTIAL.
FT	TRANSMEM 166 186				POTENTIAL.
FT	TRANSMEM 215 235				POTENTIAL.
FT	TRANSMEM 249 269				POTENTIAL.
FT	TRANSMEM 301 321				POTENTIAL.
FT	TRANSMEM 359 379				POTENTIAL.
FT	TRANSMEM 405 425				POTENTIAL.
FT	TRANSMEM 443 463				POTENTIAL.
FT	TRANSMEM 468 488				POTENTIAL.
SQ	SEQUENCE 503 AA: 54960 MW; 72E63CBC CRC32;				

**THIS PAGE BLANK (uspto)**

Best Local Similarity 90.08; Pred. No. 2.60e-01;  
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 82 APAPATSWPL 91  
||||| |||||  
QY 1 APAPATSWPL 10

RESULT 13  
ID O36006 PRELIMINARY; PRT; 391 AA.  
AC O36006;  
DT 01-JAN-1998 (TReMBLrel. 05, Created)  
DT 01-JAN-1998 (TReMBLrel. 05, Last sequence update)  
DT 01-NOV-1999 (TReMBLrel. 12, Last annotation update)  
DE CELLULAR TUMOR ANTIGEN P53.  
GN P53.  
OS Marmota monax (Woodchuck).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Rodentia; Sciurognathi; Scluridae; Sciurinae; Marmota.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 97376996.  
RA FEITELSON M.A., RANGANATHAN P.N., CLAYTON M.M., ZHANG S.M.;  
RT Partial characterization of the woodchuck tumor suppressor, p53, and  
RT its interaction with woodchuck hepatitis virus X antigen in  
RT hepatocarcinogenesis.  
RL Oncogene 15:327-336(1997).  
CC -1- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT  
CC PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL  
CC CYCLE REGULATION, IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY  
CC REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED  
CC FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF  
CC CYCLIN-DEPENDENT KINASES (BY SIMILARITY).  
CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
DR EMBL: AJ001022; CA04478.1; -.  
DR HSSP: P04637; ITS.  
DR PROSITE; PS00348; P53; 1.  
DR PFAM: PF00870; P53; 1.  
DR PRINTS; PR00386; P53SUPPRESSR.  
KW Anti-oncogene; DNA-binding; Transcription regulation; Activator;  
KW Nuclear protein; Phosphorylation.  
SQ SEQUENCE 391 AA; 4368 MW; 95FAB8F2 CRC32;

Query Match 86.58; Score 64; DB 6; Length 391;  
Best Local Similarity 80.08; Pred. No. 5.76e-01;  
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 82 APSPATSWPL 91  
||||| |||||  
QY 1 APAPATSWPL 10

RESULT 14  
ID F89002 PRELIMINARY; PRT; 378 AA.  
AC F89002;  
DT 01-MAY-1997 (TReMBLrel. 03, Created)  
DT 01-MAY-1997 (TReMBLrel. 03, Last sequence update)  
DT 01-NOV-1999 (TReMBLrel. 12, Last annotation update)  
DE P53 (FRAGMENT).  
OS Mastomys natalensis papillomavirus (Mnpv).  
OC Viruses; dsDNA viruses, no RNA stage; Papovaviridae; Papillomavirus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA LUQUE E.A., TANG L.H., MODLIN I.M.;  
RL Gastroenterology 0:0-0(0).  
DR EMBL: U48616; AAB41831.1; -.  
DR HSSP: P04637; IPET.  
DR PFAM: PF00870; P53; 1.  
FT NON\_TER 1  
SQ SEQUENCE 378 AA; 42062 MW; B4436760 CRC32;

Query Match 85.18; Score 63; DB 14; Length 378;  
Best Local Similarity 80.08; Pred. No. 8.54e-01;

Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 69 APAPATPWPL 78  
||||| |||||  
QY 1 APAPATSWPL 10

RESULT 15  
ID O70366 PRELIMINARY; PRT; 390 AA.  
AC O70366;  
DT 01-AUG-1998 (TReMBLrel. 07, Created)  
DT 01-AUG-1998 (TReMBLrel. 07, Last sequence update)  
DT 01-NOV-1999 (TReMBLrel. 12, Last annotation update)  
DE CELLULAR TUMOR ANTIGEN P53.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-LYMPHOID LEUKEMIA;  
RA FROSTESJO L., NILSSON J., WANDZIOCH E., HEBY O.;  
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT  
CC PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL  
CC CYCLE REGULATION, IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY  
CC REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED  
CC FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF  
CC CYCLIN-DEPENDENT KINASES (BY SIMILARITY).  
CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
DR EMBL: AF051368; AAC05704.1; -.  
DR HSSP: P04637; IPET.  
DR PROSITE; PS00348; P53; 1.  
DR PFAM: PF00870; P53; 1.  
DR PRINTS; PR00386; P53SUPPRESSR.  
KW Anti-oncogene; DNA-binding; Transcription regulation; Activator;  
KW Nuclear protein; Phosphorylation.  
SQ SEQUENCE 390 AA; 43430 MW; EDF4C9AA CRC32;

Query Match 85.18; Score 63; DB 11; Length 390;  
Best Local Similarity 80.08; Pred. No. 8.54e-01;  
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 81 APAPATPWPL 90  
||||| |||||  
QY 1 APAPATSWPL 10

Search completed: Sat Apr 15 00:24:20 2000  
Job time : 92 secs.



DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
 DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)  
 DE CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).  
 GN P53.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 92007731.  
 RA FARRELL P.J., ALLAN G., SHANAHAN F., VOUSDEN K.H., CROOK T.;  
 RT "p53 is frequently mutated in Burkitt's lymphoma cell lines."  
 RL EMO J. 10:2879-2887(1991).  
 CC -|- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT  
 CC PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL  
 CC CYCLE REGULATION, IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY  
 CC REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED  
 CC FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF  
 CC CYCLIN-DEPENDENT KINASES (BY SIMILARITY).  
 CC -|- SUBCELLULAR LOCATION: NUCLEAR.  
 DR EMBL; X60018; CAA42633.1; -.  
 DR HSSP; P04637; 1SAH.  
 DR PFAM; PS00348; P53; 1.  
 DR PROSITE; PS00870; P53; 1.  
 DR Anti-oncogene; DNA-binding; Transcription regulation; Activator;  
 KW Nuclear protein; Phosphorylation.  
 FT VARIANT 163 163 H -> Y.  
 FT NON\_TER 393 393  
 SQ SEQUENCE 393 AA; 43627 MW; AFD8A9E3 CRC32;

Query Match 100.0%; Score 74; DB 4; Length 393;  
 Best Local Similarity 100.0%; Pred. No. 9.71e-03;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 84 APAPAPSWPL 93  
 |||||  
 QY 1 APAPAPSWPL 10

RESULT 7  
 ID Q16535 PRELIMINARY; PRT; 393 AA.  
 AC Q16535;  
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
 DT 01-NOV-1999 (TrEMBLrel. 01, Last sequence update)  
 DE P53 TRANSFORMATION SUPPRESSOR (FRAGMENT).  
 GN P53.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 92007731.  
 RA FARRELL P.J., ALLAN G., SHANAHAN F., VOUSDEN K.H., CROOK T.;  
 RT "p53 is frequently mutated in Burkitt's lymphoma cell lines."  
 RL EMO J. 10:2879-2887(1991).  
 DR EMBL; X60017; CAA42632.1; -.  
 DR EMBL; X60015; CAA42630.1; -.  
 DR HSSP; P04637; 1SAH.  
 DR PFAM; PS00870; P53; 1.  
 DR PROSITE; PS00870; P53; 1.  
 FT VARIANT 248 248 Q -> R.  
 FT NON\_TER 393 393  
 SQ SEQUENCE 393 AA; 43684 MW; 239818A9 CRC32;

Query Match 100.0%; Score 74; DB 4; Length 393;  
 Best Local Similarity 100.0%; Pred. No. 9.71e-03;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 84 APAPAPSWPL 93  
 |||||  
 QY 1 APAPAPSWPL 10

RESULT 8  
 ID Q16809 PRELIMINARY; PRT; 393 AA.  
 AC Q16809;  
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
 DT 01-NOV-1999 (TrEMBLrel. 01, Last sequence update)  
 DE CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).  
 GN P53.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 92007731.  
 RA FARRELL P.J., ALLAN G., SHANAHAN F., VOUSDEN K.H., CROOK T.;  
 RT "p53 is frequently mutated in Burkitt's lymphoma cell lines."  
 RL EMO J. 10:2879-2887(1991).  
 CC -|- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT  
 CC PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL  
 CC CYCLE REGULATION, IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY  
 CC REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED  
 CC FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF  
 CC CYCLIN-DEPENDENT KINASES (BY SIMILARITY).  
 CC -|- SUBCELLULAR LOCATION: NUCLEAR.  
 DR EMBL; X60019; CAA42634.1; -.  
 DR HSSP; P04637; 1SAH.  
 DR PROSITE; PS00348; P53; 1.  
 DR PFAM; PF00870; P53; 1.  
 DR Anti-oncogene; DNA-binding; Transcription regulation; Activator;  
 KW Nuclear protein; Phosphorylation.  
 FT VARIANT 213 213 Q -> R.  
 FT NON\_TER 393 393  
 SQ SEQUENCE 393 AA; 43684 MW; CB70BD7F CRC32;

Query Match 100.0%; Score 74; DB 4; Length 393;  
 Best Local Similarity 100.0%; Pred. No. 9.71e-03;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 84 APAPAPSWPL 93  
 |||||  
 QY 1 APAPAPSWPL 10

RESULT 9  
 ID Q16848 PRELIMINARY; PRT; 393 AA.  
 AC Q16848;  
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
 DT 01-NOV-1999 (TrEMBLrel. 01, Last sequence update)  
 DE CELLULAR TUMOR ANTIGEN P53.  
 GN TP53.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 87089826.  
 RA HARRIS N., BRILL E., SHOHAT O., PROKOCIMER M., WOLF D., ARAI N.,  
 RA ROTTER V.;  
 RT "Molecular basis for heterogeneity of the human p53 protein."  
 RL Mol. Cell. Biol. 6:4650-4656(1986).  
 CC -|- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT  
 CC PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL  
 CC CYCLE REGULATION, IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY  
 CC REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED  
 CC FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF  
 CC CYCLIN-DEPENDENT KINASES (BY SIMILARITY).  
 CC -|- SUBCELLULAR LOCATION: NUCLEAR.  
 DR EMBL; M14694; AAA61211.1; -.  
 DR HSSP; P04637; 1TSR.  
 DR PROSITE; PS00348; P53; 1.  
 DR PFAM; PF00870; P53; 1.  
 DR PRINTS; PR00386; P53SUPPRESSR.



```

RN  SEQUENCE FROM N.A.
RP  MEDLINE; 92007731.
RA  FARRELL P.J., ALLAN G., SHANAHAN F., VOUSDEN K.H., CROOK T.;
RT  "p53 is frequently mutated in Burkitt's lymphoma cell lines.";
RL  EMBO J. 10:2879-2887(1991).
DR  EMBL; X60016; CAA42631.1; -.
DR  HSSP; P04637; 1SAH.
DR  PFAM; PF00870; P53; 1.
FT  VARIANT 238 Y -> C.
FT  NON_TER 393
SQ  SEQUENCE 393 AA; 43713 MW; A01E1523 CRC32;

Query Match 100.0%; Score 74; DB 4; Length 393;
Best Local Similarity 100.0%; Pred. No. 9.71e-03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 84 APAPAPSWPL 93
QY 1 APAPAPSWPL 10

RESULT 3
ID Q15086 PRELIMINARY; PRT; 393 AA.
AC Q15086;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
DE P53 TRANSFORMATION SUPPRESSOR (FRAGMENT).
GN P53.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 92007731.
RA FARRELL P.J., ALLAN G., SHANAHAN F., VOUSDEN K.H., CROOK T.;
RT "p53 is frequently mutated in Burkitt's lymphoma cell lines.";
RL EMBO J. 10:2879-2887(1991).
DR EMBL; X60013; CAA42628.1; -.
DR HSSP; P04637; 1SAH.
DR PFAM; PF00870; P53; 1.
FT VARIANT 246 T -> M.
FT NON_TER 393
SQ SEQUENCE 393 AA; 43682 MW; 943B62A3 CRC32;

Query Match 100.0%; Score 74; DB 4; Length 393;
Best Local Similarity 100.0%; Pred. No. 9.71e-03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 84 APAPAPSWPL 93
QY 1 APAPAPSWPL 10

RESULT 4
ID Q16810 PRELIMINARY; PRT; 393 AA.
AC Q16810;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
DE CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).
GN P53.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 92007731.
RA FARRELL P.J., ALLAN G., SHANAHAN F., VOUSDEN K.H., CROOK T.;
RT "p53 is frequently mutated in Burkitt's lymphoma cell lines.";
RL EMBO J. 10:2879-2887(1991).
DR EMBL; X60016; CAA42631.1; -.
DR HSSP; P04637; 1SAH.
DR PFAM; PF00870; P53; 1.
FT VARIANT 238 Y -> C.
FT NON_TER 393
SQ SEQUENCE 393 AA; 43713 MW; A01E1523 CRC32;

Query Match 100.0%; Score 74; DB 4; Length 393;
Best Local Similarity 100.0%; Pred. No. 9.71e-03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 84 APAPAPSWPL 93
QY 1 APAPAPSWPL 10

RESULT 5
ID Q16807 PRELIMINARY; PRT; 393 AA.
AC Q16807;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
DE CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).
GN P53.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 92007731.
RA FARRELL P.J., ALLAN G., SHANAHAN F., VOUSDEN K.H., CROOK T.;
RT "p53 is frequently mutated in Burkitt's lymphoma cell lines.";
RL EMBO J. 10:2879-2887(1991).
DR EMBL; X60013; CAA42628.1; -.
DR HSSP; P04637; 1SAH.
DR PFAM; PF00870; P53; 1.
FT VARIANT 246 T -> M.
FT NON_TER 393
SQ SEQUENCE 393 AA; 43682 MW; 943B62A3 CRC32;

Query Match 100.0%; Score 74; DB 4; Length 393;
Best Local Similarity 100.0%; Pred. No. 9.71e-03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 84 APAPAPSWPL 93
QY 1 APAPAPSWPL 10

RESULT 6
ID Q16808 PRELIMINARY; PRT; 393 AA.
AC Q16808;
DT 01-NOV-1996 (TREMBlrel. 01, Created)

```

\*\*\*\*\*

W P S R L H (TM)

\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Sat Apr 15 00:22:48 2000; MasPar time 7.23 Seconds  
Tabular output not generated. 95.909 Million cell updates/sec

Title: >US-08-452-843-14  
Description: (1-10) from US08452843.pep  
Perfect Score: 74  
Sequence: 1 APAPAPSWPL 10

Scoring table: PAM 150  
Gap 15

Searched: 225878 seqs, 69334122 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: Sptrembl12  
1:sp.archaea 2:sp.bacteria 3:sp.fungi 4:sp.human  
5:sp.invertebrate 6:sp.mammal 7:sp.mhc 8:sp.organelle  
9:sp.phage 10:sp.plant 11:sp.rodent 12:sp.unclassified  
13:sp.vertebrate 14:sp.virus

Statistics: Mean 23.348; Variance 39.816; scale 0.586

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	74	100.0	393	4	Q15087 P53 TRANSFORMATION SUP	9.71e-03
2	74	100.0	393	4	Q15088 P53 TRANSFORMATION SUP	9.71e-03
3	74	100.0	393	4	Q15086 P53 TRANSFORMATION SUP	9.71e-03
4	74	100.0	393	4	Q15810 CELLULAR TUMOR ANTIGEN	9.71e-03
5	74	100.0	393	4	Q16807 CELLULAR TUMOR ANTIGEN	9.71e-03
6	74	100.0	393	4	Q16808 CELLULAR TUMOR ANTIGEN	9.71e-03
7	74	100.0	393	4	Q16535 P53 TRANSFORMATION SUP	9.71e-03
8	74	100.0	393	4	Q16809 CELLULAR TUMOR ANTIGEN	9.71e-03
9	74	100.0	393	4	Q16848 CELLULAR TUMOR ANTIGEN	9.71e-03
10	74	100.0	393	4	Q16811 CELLULAR TUMOR ANTIGEN	9.71e-03
11	71	95.9	285	6	Q95326 CELLULAR TUMOR ANTIGEN	3.39e-02
12	66	89.2	391	11	Q9WUR6 CELLULAR TUMOR ANTIGEN	2.60e-01
13	64	86.5	391	6	Q36006 CELLULAR TUMOR ANTIGEN	5.76e-01
14	63	85.1	378	14	P99002 P53 (FRAGMENT)	8.54e-01
15	63	85.1	390	11	Q70366 CELLULAR TUMOR ANTIGEN	8.54e-01
16	61	82.4	1049	4	Q94937 KIAA0881 PROTEIN.	1.86e-00
17	58	78.4	1615	4	L1P0P0TEIN RECEPTOR R	5.84e-00
18	57	77.0	658	10	Q05214 CYSTEINE PROTEASE.	8.50e-00
19	56	75.7	781	10	T14N5.13 PROTEIN.	1.23e-01
20	56	75.7	837	1	ATP-DEPENDENT RNA HELI	1.23e+01

21	56	75.7	2554	5	Q24512 SEVENLESS PROTEIN (EC	1.23e+01
22	55	74.3	436	2	O54450 METHYLAMMONIUM TRANSPO	1.78e+01
23	55	74.3	814	4	MDCL5	1.78e+01
24	55	74.3	814	4	Q13493 METARGIDIN PRECURSOR.	1.78e+01
25	55	74.3	1937	2	O30482 PKS MODULE 4.	1.78e+01
26	54	73.0	314	4	O9Y6M0 TESTISIN.	2.56e+01
27	54	73.0	441	2	O87855 PUTATIVE TRANSCRIPTION	2.56e+01
28	54	73.0	512	10	O49396 CYTOCHROME P450-LIKE P	2.56e+01
29	54	73.0	523	10	O49394 CYTOCHROME P450-LIKE P	2.56e+01
30	54	73.0	600	2	Q9ZG01 YBP.	2.56e+01
31	53	71.6	206	14	Q79690 NEF PROTEIN.	3.67e+01
32	53	71.6	206	14	Q79689 NEF PROTEIN.	3.67e+01
33	53	71.6	206	14	Q72427 NEF PROTEIN.	3.67e+01
34	53	71.6	206	14	Q72426 NEF PROTEIN.	3.67e+01
35	53	71.6	206	14	Q79691 NEF PROTEIN.	3.67e+01
36	53	71.6	206	14	Q79706 NEF PROTEIN.	3.67e+01
37	53	71.6	206	14	Q79707 NEF PROTEIN.	3.67e+01
38	53	71.6	206	14	Q79685 NEF PROTEIN.	3.67e+01
39	53	71.6	206	14	Q79708 NEF PROTEIN.	3.67e+01
40	53	71.6	206	14	Q79710 NEF PROTEIN.	3.67e+01
41	53	71.6	206	14	Q72425 NEF PROTEIN (FRAGMENT)	3.67e+01
42	53	71.6	206	14	Q79701 NEF PROTEIN.	3.67e+01
43	53	71.6	206	14	Q79702 NEF PROTEIN.	3.67e+01
44	53	71.6	206	14	Q79703 NEF PROTEIN.	3.67e+01
45	53	71.6	207	14	Q79725 NEF PROTEIN (FRAGMENT)	3.67e+01

## ALIGNMENTS

RESULT	ID	Q15087	PRELIMINARY;	PRT;	393 AA.
AC	Q15087;				
DT	01-NOV-1996	(Tremblrel. 01, Created)			
DT	01-NOV-1996	(Tremblrel. 01, Last sequence update)			
DT	01-NOV-1999	(Tremblrel. 12, Last annotation update)			
DE	P53	TRANSFORMATION SUPPRESSOR (FRAGMENT).			
GN	P53.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;				
OC	Eutheria; Primates; Catarrhini; Hominidae; Homo.				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE; 92007731.				
RA	FARRELL P.J., ALLAN G., SHANAHAN F., VOUSDEN K.H., CROOK T.;				
RT	"p53 is frequently mutated in Burkitt's lymphoma cell lines."				
RL	EMBO J. 10:2879-2887(1991).				
DR	EMBL; X60014; CA442629.1; -.				
DR	HSSP; P04637; 1SAH.				
DR	PFAM; PF00870; P53; 1.				
FT	VARIANT 237 237				
FT	NON-TER 393 393				
FT	SEQUENCE 393 AA; 43694 MW; 9BB81992 CRC32;				
FT	I -> M.				

Query Match 100.0%; Score 74; DB 4; Length 393;  
Best Local Similarity 100.0%; Pred. No. 9.71e-03;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 84 APAPAPSWPL 93

QY 1 APAPAPSWPL 10

RESULT 2 PRELIMINARY; PRT; 393 AA.

ID	Q15088				
AC	Q15088;				
DT	01-NOV-1996	(Tremblrel. 01, Created)			
DT	01-NOV-1996	(Tremblrel. 01, Last sequence update)			
DT	01-NOV-1999	(Tremblrel. 12, Last annotation update)			
DE	P53	TRANSFORMATION SUPPRESSOR (FRAGMENT).			
GN	P53.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;				
OC	Eutheria; Primates; Catarrhini; Hominidae; Homo.				

THIS PAGE BLANK (USPTO)

RESULT 14  
ID R10681 standard; Protein; 589 AA.  
AC R10681:  
DT 17-APR-1991 (first entry)  
DE Polyhydroxybutyrate polymerase enzyme.  
KW Polyester biopolymers; polyhydroxybutyrate; polyhydroxy alkanoate;  
OS beta-ketothiolase; acetoacetyl CoA reductase.  
PN W09100917-A.  
PD 24-JAN-1991.  
PF 10-JUL-1990; U03851.  
PR 10-JUL-1989; US-378155.  
PA (MASI ) MASSACHUSETTS INST TECH.  
PI Peoples Op, Sinskey AJ;  
DR WPI: 91-051341/07.  
DR N-PSDB: Q10502.  
PT Construction and modification of polyester bio:polymers - by  
PT introduction of poly-hydroxy-butyrate and -alkanoate genes into  
PT bacteria or plants  
PS Disclosure; fig 4; 64pp; English.  
CC This Alcaligenes eutrophus polyhydroxybutyrate (PHB) polymerase enzyme  
CC is essential to the biosynthesis of PHB. The use of recombinant methods  
CC for producing such enzymes, required for polyester biopolymer synthesis,  
CC allows for the control and modification of the synthesis process.  
CC See also Q10499-501 and  
CC Q10503.  
SQ Sequence 589 AA;

Query Match 69.4%; Score 50; DB 1; Length 589;  
Best Local Similarity 55.6%; Pred. No. 2.50e+02;  
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;  
  
DB 111 LPYRFAAAF 119  
QY :|:::|||||  
1 PPFKYAAAF 9

RESULT 15  
ID R71325 standard; Protein; 672 AA.  
AC R71325;  
DT 21-OCT-1995 (first entry)  
DE Poly-beta-hydroxyalkanoate-synthase.  
KW Poly-beta-hydroxyalkanoate-synthase; transgenic plant;  
KW poly-beta-hydroxyalkanoate; poly-beta-hydroxybutyrate;  
KW biodegradable thermoplastic.  
OS Alcaligenes eutrophus.  
PN W09505472-A.  
PD 23-FEB-1995.  
PF 17-AUG-1994; U09265.  
PR 17-AUG-1993; US-108193.  
PR 06-JUN-1994; US-254357.  
PA (UNMS ) UNIV MICHIGAN STATE.  
PI Nawrath C, Poirier Y, Somerville CR;  
DR WPI: 95-098770/13.  
DR N-PSDB: R71325.  
PT Transgenic plant material with plastid(s) contg. the enzymes for  
PT synthesis of poly:hydroxy:alkanoate(s) - express  
PT poly:hydroxy:butyrate and have good growth and seed formation.  
PS Claim 2; Page 62-64; 88pp; English.  
CC The poly-beta-hydroxyalkanoate-synthase gene (phbC) from A.  
CC eutrophus is cloned under the control of an Arabidopsis thaliana  
CC seed storage protein promoter for plastid tissue-specific  
CC gene expression in a transgenic plant. When expressed with the  
CC 3-ketothiolase (phbA) and acetyl-CoA-reductase (phbB) genes, a  
CC poly-beta-hydroxyalkanoate (PHA), specifically poly-beta-  
CC hydroxybutyrate (PHB), is expressed in the transgenic plant  
CC (preferably a Brassica e.g. rape). PHB and related PHAs are  
CC biodegradable thermoplastics with many useful applications.  
SQ Sequence 672 AA;

Query Match 69.4%; Score 50; DB 1; Length 672;  
Best Local Similarity 55.6%; Pred. No. 2.50e+02;

Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;  
  
DB 194 LPYRFAAAF 202  
QY :|:::|||||  
1 PPFKYAAAF 9

Search completed: Fri Apr 14 23:00:20 2000  
Job time : 38 secs.

PT modulators useful in therapy  
 PS Claim 11; Fig 2; 52pp; English.  
 CC The present sequence is a novel human protein that functions as a  
 CC ras carboxyl terminal processing protein, and optionally as a ras  
 CC processing enzyme (i.e. ras protease). cDNA (see X24921) encoding  
 CC the protein was cloned from a human colorectal adenocarcinoma cDNA  
 CC library following a homology search using yeast RCE1, the protease  
 CC responsible for ras processing in yeast. The cDNA has been  
 CC expressed in recombinant host cells which produce active  
 CC recombinant protein. The recombinant protein, and recombinant  
 CC host cells are utilised in a method for identifying modulators  
 CC of the enzyme activity, useful for treating a condition mediated  
 CC by activated ras protein. Inhibition of the human ras protease  
 CC should be efficacious for cancer treatment.  
 SQ Sequence 338 AA;

Query Match 69.4%; Score 50; DB 1; Length 338;  
 Best Local Similarity 55.6%; Pred. No. 2.50e+02;  
 Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Db 241 FQSYTAVF 249  
 | | | | |  
 QY 1 FPFKYAAAF 9

RESULT 11  
 ID W20801 standard; protein; 496 AA.

AC W20801;  
 DE 16-JUL-1997 (first entry)  
 DE H. pylori inner membrane protein, O9ap11406orf15.  
 KW Cytoplasmic; vaccine; prevention; treatment; infection; identification;  
 KW binding compound; bacterium; life cycle; activator; bacteria; inhibitor;  
 KW duodenal ulcer disease; chronic gastritis; diagnosis; envelope.  
 OS Helicobacter pylori.  
 PN W09640893-A1.  
 PD 19-DEC-1996.  
 PF 06-JUN-1996; U09122.  
 PR 07-JUN-1995; US-487032.  
 PR 01-APR-1996; US-630405.  
 PA (ASTR) ASTRA AB.  
 PI Berglindh OT, Smith D, Mellgaard BL;  
 DR WPI: 97-052306/05.  
 DR N-PSDB: T68054.

PT Helicobacter pylori nucleic acid sequences and related  
 PT polypeptide(s) - useful for vaccines to treat or prevent H. pylori  
 PT infection, and to detect Helicobacter  
 PS Claim 56; Page 1207-1208; 1481pp; English.  
 CC The present sequence is a H. pylori inner membrane protein.  
 CC The protein may be used in a vaccine to prevent or treat H. pylori  
 CC infection or to identify H. pylori polypeptide binding compounds,  
 CC useful as potential H. pylori life cycle activators or inhibitors.  
 CC The genomic sequence of H. pylori (ATCC 55679) was determined from  
 CC overlapping contigs generated by mechanically shearing the bacterial  
 CC DNA. The sequences were analysed for ORF of at least 180 nucleotides,  
 CC and the predicted coding regions defined by computer evaluation. To  
 CC identify likely H. pylori antigens for vaccine development, the amino  
 CC acid sequences predicted from various ORF were analysed for significant  
 CC homology to other known or exported membrane proteins. Having identified  
 CC and determined the sequences of interest, particular regions can be  
 CC isolated from H. pylori by PCR amplification for recombinant polypeptide  
 CC production, e.g. in E. coli hosts.  
 SQ Sequence 496 AA;

Query Match 69.4%; Score 50; DB 1; Length 496;  
 Best Local Similarity 55.6%; Pred. No. 2.50e+02;  
 Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 343 FARTYICAF 351  
 | | | | |  
 QY 1 FPFKYAAAF 9

RESULT 12

ID W75421 standard; protein; 511 AA.  
 AC W75421;  
 DT 16-MAR-1999 (first entry)  
 DE T.thermophilus nitrate reductase beta subunit.  
 KW Heat-stable; nitrate reductase; temperature; detection; food; toxicity;  
 KW carcinogen.  
 OS Thermus thermophilus.  
 PN ES2121561-A1.  
 PD 16-NOV-1998.  
 PF 09-MAY-1997; 001003.  
 PR 09-MAY-1997; ES-001003.  
 PA (UYMA-) UNIV AUTONOMA MADRID.  
 DR WPI: 99-001909/01.  
 DR Heat stable nitrate reductase for high temperature nitrate detection  
 PT - comprises Thermus thermophilus derivative enhancing nitrite or  
 PT nitrate reduction  
 PS Disclosure; Fig 3; 8pp; Spanish.  
 CC This sequence represents the amino acid sequence of the Thermus  
 CC thermophilus heat-stable nitrate reductase beta subunit. Heat stable  
 CC nitrate reductase can be used for high-temperature detection of nitrates  
 CC in samples, e.g. in food, where high levels of nitrates can be toxic or  
 CC carcinogenic.  
 SQ Sequence 511 AA;

Query Match 69.4%; Score 50; DB 1; Length 511;  
 Best Local Similarity 55.6%; Pred. No. 2.50e+02;  
 Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Db 103 FTFRYADLF 111  
 | | | | |  
 QY 1 FPFKYAAAF 9

RESULT 13

ID R32190 standard; protein; 589 AA.  
 AC R32190;  
 DT 30-MAY-1993 (first entry)  
 DE Sequence encoded by the PHB synthase (phbC) gene of the  
 DE polyhydroxybutyrate (PHB) operon  
 KW Operon; polyhydroxyalkanoate; polyhydroxybutyrate synthase.  
 OS Alcaligenes eutrophus.  
 PN W09302187-A.  
 PD 04-FEB-1993.  
 PF 13-JUL-1992; U05786.  
 PR 19-JUL-1991; US-732243.  
 PA (UYMA-) UNIV MADISON JAMES.  
 PA (UNMS) UNIV MICHIGAN STATE.  
 PI Dennis DE, Poirier Y, Somerville CR;  
 DR WPI: 93-058785/07.  
 DR N-PSDB: Q36660.  
 PT Transgenic plants producing poly:hydroxy-alkanoate polymer(s) -  
 PT obtd. by transformation with DNA encoding 3-ketothiolase,  
 PT acetoacetyl-CoA reductase and PHA synthase  
 PS Disclosure; Fig 2; 70pp; English.  
 CC The nucleotide sequence of the PHB operon was obtained from Janes, B.  
 CC Holiar, J. and Dennis, D. in Dawes, E.A. (ed.) Novel Biodegradable  
 CC Polymers, Kluwer Academic Publishers, 175-190 (1990). It contains  
 CC the genes from PHB synthase, 3-ketothiolase and acetoacetyl-CoA  
 CC reductase. The inventors claim a transgenic plant material contg.  
 CC foreign DNA encoding a peptide which exhibits 3-ketothiolase activity,  
 CC pref. where the DNA is an open reading from between nucleotides  
 CC 2696-3877 (phb A gene), 842-2611 (phb C gene) or 3952-4692 (phb B  
 CC gene) of the Alcaligenes eutrophus PHB operon.  
 SQ Sequence 589 AA;

Query Match 69.4%; Score 50; DB 1; Length 589;  
 Best Local Similarity 55.6%; Pred. No. 2.50e+02;  
 Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 111 LPYRFAAAF 119  
 | | | | |  
 QY 1 FPFKYAAAF 9

CC The genomic sequence of *H. pylori* (ATCC 55679) was determined from  
 CC overlapping contigs generated by mechanically shearing the bacterial  
 CC DNA. The sequences were analysed for ORF of at least 180 nucleotides,  
 CC and the predicted coding regions defined by computer evaluation. To  
 CC identify likely *H. pylori* antigens for vaccine development, the amino  
 CC acid sequences predicted from various ORF were analysed for significant  
 CC homology to other known or exported membrane proteins. Having identified  
 CC and determined the sequences of interest, particular regions can be  
 CC isolated from *H. pylori* by PCR amplification for recombinant polypeptide  
 CC production, e.g. in *E. coli* hosts.  
 SQ Sequence 326 AA;

Query Match 69.4%; Score 50; DB 1; Length 326;  
 Best Local Similarity 55.6%; Pred. No. 2.50e+02;  
 Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 191 FAFYIGAF 199

QY 1 FPFYIAAF 9

RESULT 8  
 ID W24686 standard; Protein; 326 AA.

AC W24686; 1997 (first entry)

DE *H. pylori* inner membrane protein 6093906.aa.

KW Transmembrane; cytoplasmic; cell envelope; flagella; transport;

KW secreted; periplasmic; chronic gastritis; duodenal ulcer disease;

KW activator; inhibitor; bacterial life cycle; vaccine; immunise;

KW detection; antisense; inhibition.

OS Helicobacter pylori.

FH Key Location/Qualifiers

FT ms\_difference 6 /note="encoded by ARC"

FT W09719098-A1.

PD 29-MAY-1997.

PF 15-NOV-1996; U18542.

PR 17-NOV-1995; US-561469.

PA (ASTR ) ASTRA AB.

PI Smith DH;

DR WPI: 97-298052/27.

DR N-PSDB; T77504.

PT Helicobacter pylori nucleic acid sequences and related proteins -

PT used for diagnostics and therapeutics

PS Claim 18; Page 193; 235pp; English.

CC This sequence represents an *H. pylori* inner membrane protein.

CC Helicobacter pylori has been strongly linked to chronic gastritis and

CC duodenal ulcer disease. The nucleic acid sequences of the invention

CC are used to evaluate compounds, especially activators or inhibitors of

CC bacterial life cycle, for the ability to bind an *H. pylori* nucleic acid

CC sequence. The nucleic acid sequences, and corresponding proteins, are

CC also useful for generating vaccines for immunising subjects against *H.*

CC *pylori* or for use in detecting the presence of Helicobacter species in

CC a sample. Antisense nucleic acid sequences of these sequences are

CC used to inhibit expression of a gene from Helicobacter species. *H.*

CC *pylori* whole genomic DNA was isolated and nebulised to a median size of

CC 2000 bp. Purified DNA fragments were blunt-ended and ligated to unique

CC BstXI-linker adapters in 100-1000 fold molar excess. These linkers are

CC complementary to the BstXI-cut PMPX vectors, while the overhang is not

CC self-complementary. Therefore the linkers will not concatamerise nor

CC will the cut vector re-ligate itself easily. The linker-adaptor inserts

CC were ligated to each of the 20 PMPX vectors to construct a series of

CC shotgun subclone libraries. The purified DNA samples were then

CC sequenced.

CC Note: The ORF/protein reference number for this sequence was obtained

CC from the related specification, W09640893.

SQ Sequence 326 AA;

Query Match 69.4%; Score 50; DB 1; Length 326;

Best Local Similarity 55.6%; Pred. No. 2.50e+02;

Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 191 FAFYIGAF 199

QY 1 FPFYIAAF 9

RESULT 9

ID W86010 standard; Protein; 329 AA.

AC W86010; 1999 (first entry)

DE Mouse CAAX processing enzyme RCE1 homologue mRCElp.

KW RCE1; RCElp; mRCElp; CAAX processing enzyme; mouse; tumour; cancer;

KW therapy; diagnosis; Ras protein; endoproteinase.

OS Mus sp.

PN W09854333-A2.

PD 03-DEC-1998.

PF 02-JUN-1998; U11415.

PR 14-JUL-1997; US-052389.

PR 02-JUN-1997; US-047369.

PA (ACAC-) ACACIA BIOSCIENCES INC.

PI Ashby MN, Dimster-Denk DG, Phillips JW;

DR WPI: 99-059843/05.

DR N-PSDB; V80323-24.

PT New DNA encoding mammalian CAAX-processing enzymes - used e.g. to

PT treat CAAX-protein mediated diseases such as cancers and tumours

PT associated with mutant Ras

PS Claim 15; Fig 2B; 98pp; English.

CC This is the amino acid sequence of mRCElp, a murine functional

CC homologue of the yeast prenylation-dependent CAAX endoproteinase

CC RCElp that contributes to the processing of a-factor and the yeast

CC Ras protein. This mammalian homologue represents a potential

CC target to block the oncogenic action of mutant Ras protein in

CC tumours or, more generally, to modulate the activity of prenylated

CC peripheral membrane proteins. The mRCElp amino acid sequence was

CC deduced from the nucleotide sequence of isolated cDNA and genomic

CC clones (see V80323-24). Host cells transfected with mammalian

CC CAAX processing enzyme DNA (see V80322-25) can be used to produce

CC recombinant polypeptides (see W86009-12) used for in vitro

CC screening of inhibitors and to raise antibodies. The inhibitors

CC are used to treat CAAX-protein mediated diseases, especially

CC cancers and tumours associated with abnormal Ras activity.

CC Antibodies are used to screen for expression of CAAX processing

CC proteins, for affinity purification and in immunoassays to

CC determine levels of CAAX processing proteins or their subcellular

CC localisation and to confirm interaction with candidate binding

CC proteins.

SQ Sequence 329 AA;

Query Match 69.4%; Score 50; DB 1; Length 329;

Best Local Similarity 55.6%; Pred. No. 2.50e+02;

Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Db 232 FQFSYTAVF 240

QY 1 FPFYIAAF 9

RESULT 10

ID W98105 standard; Protein; 338 AA.

AC W98105;

DT 05-JUL-1999 (first entry)

DE Guman ras carboxy-terminal processing protein.

KW Ras carboxy-terminus processing protein; protease; human; cancer;

KW therapy.

OS Homo sapiens.

PN W09914343-A1.

PD 25-MAR-1999.

PF 18-SEP-1998; U19746.

PR 19-SEP-1997; US-059401.

PA (ORTH ) ORTHO-MCNEIL PHARM INC.

PI Chamberlain H, Farrell F, Galindo J, Huvar A, Johnson D,

PI Jolliffe L, Patel L;

DR WPI: 99-229542/19.

DR N-PSDB; X24921.

PT New ras carboxyl terminal processing protein useful for identifying

CC They can also be used to develop products for diagnosis and therapy.  
CC The proteins obtained may have cytokine activity, cell  
CC proliferation/differentiation activity, haematopoiesis regulating  
CC activity, tissue growth regulating activity, reproductive hormone  
CC regulating activity, chemotactic/chemokinetic activity, haemostatic and  
CC thrombolytic activity, receptor/ligand activity, anti-inflammatory  
CC activity, tumour inhibition activity or other activities. The products  
CC can be used in forensic, gene therapy and chromosome mapping procedures.  
CC The sequences can also be used for obtaining corresponding promoter  
CC sequences. The nucleic acids encoding the signal peptide can be used  
CC for directing extracellular secretion of a polypeptide or the insertion  
CC of a polypeptide into a membrane, or importing a polypeptide into  
CC a cell.  
SQ Sequence 62 AA;

Query Match 69.4%; Score 50; DB 1; Length 62;  
Best Local Similarity 44.4%; Pred. No. 2.50e+02;  
Matches 4; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 14 FPFSSQTF 22  
|||:::  
QY 1 FPFKYAAAF 9

RESULT 5  
ID Y06958 standard; Protein; 271 AA.  
AC Y06958;  
DT 05-JUL-1999 (first entry)  
DE E. chaffeensis OMP-12 protein.  
KW Outer membrane protein; OMP; Ehrlichia chaffeensis; E. canis; P30;  
KW detection; dog.  
OS Ehrlichia chaffeensis.  
PN W09913720-A1.  
PD 25-MAR-1999.  
PF 18-SEP-1998; U19600.  
PR 19-SEP-1997; US-059353.  
PA (OHIS ) UNIV OHIO STATE.  
PI Ohashi N, Rikihisa Y;  
DR WPI; 99-254290/21.  
DR N-PSDB; X34758.  
PT Novel outer membrane proteins from Ehrlichia chaffeensis and  
PT Ehrlichia canis  
PS Disclosure; Fig 18B; 55pp; English.  
CC The invention provides isolated outer membrane proteins (OMP) from  
CC Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part  
CC of the OMP family and consist of proteins OMP-1, -1(B to Z) shown  
CC in Y06943-958. The E. canis proteins form part of the P30 family and  
CC consist of proteins shown in Y06959-970. The proteins and genes are used  
CC to detect E. chaffeensis in patients and E. canis in dogs.  
SQ Sequence 271 AA;

Query Match 69.4%; Score 50; DB 1; Length 271;  
Best Local Similarity 55.6%; Pred. No. 2.50e+02;  
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 32 FPFKYSSSF 40  
|||:::  
QY 1 FPFKYAAAF 9

RESULT 6  
ID W86009 standard; Protein; 293 AA.  
AC W86009;  
DT 29-MAR-1999 (first entry)  
DE Human CAAX processing enzyme RCE1 homologue hrCelp.  
KW RCE1; Rcelp; hrCelp; CAAX processing enzyme; human; tumour;  
KW cancer; therapy; diagnosis; Ras protein; endoproteinase.  
OS Homo sapiens.  
FH Key Location/Qualifiers  
FT Region 172..175  
FT /note= "histidine-rich sequence, possible involved  
FT in substrate binding and/or catalysis"  
PN W09854333-A2.

PD 03-DEC-1998.  
PF 02-JUN-1998; U11415.  
PR 14-JUL-1997; US-052389.  
PR 02-JUN-1997; US-047369.  
PA (ACAC-) ACACIA BIOSCIENCES INC.  
PI Ashby MN Dimster-Denk DG, Phillips JW;  
DR WPI; 99-059843/05.  
DR N-PSDB; W80322.  
PT New DNA encoding mammalian CAAX-processing enzymes - used e.g. to  
PT treat CAAX-protein mediated diseases such as cancers and tumours  
PT associated with mutant Ras  
PS Claim 15; Fig 2A; 98pp; English.  
CC This is the amino acid sequence of hrCelp, a human functional  
CC homologue of the yeast prenylation-dependent CAAX endoproteinase  
CC Rcelp that contributes to the processing of a-factor and the yeast  
CC Ras protein. This mammalian homologue represents a potential  
CC target to block the oncogenic action of mutant Ras protein in  
CC tumours or, more generally, to modulate the activity of prenylated  
CC peripheral membrane proteins. The hrCelp amino acid sequence was  
CC deduced from the nucleotide sequence of cDNA clones (see V80322)  
CC isolated from a 9-wk foetus cDNA library. Host cells transformed  
CC with mammalian CAAX processing enzyme DNA sequences (see V80322-25)  
CC can be used to produce recombinant polypeptides (see W85009-12)  
CC used for in vitro screening of inhibitors and to raise antibodies.  
CC The inhibitors are used to treat CAAX-protein mediated diseases,  
CC especially cancers and tumours associated with abnormal Ras  
CC activity. Antibodies are used to screen for expression of  
CC CAAX processing proteins, for affinity purification and in  
CC immunoassays to determine levels of CAAX processing proteins or  
CC their subcellular localisation and to confirm interaction with  
CC candidate binding proteins.  
SQ Sequence 293 AA;

Query Match 69.4%; Score 50; DB 1; Length 293;  
Best Local Similarity 55.6%; Pred. No. 2.50e+02;  
Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Db 196 FQFSYAVF 204  
|||:::  
QY 1 FPFKYAAAF 9

RESULT 7  
ID W20553 standard; Protein; 326 AA.  
AC W20553;  
DT 04-AUG-1997 (first entry)  
DE H. pylori cell envelope inner membrane protein 6093906.aa.  
KW Cytoplasmic; vaccine; prevention; treatment; infection; envelope;  
KW identification; binding compound; bacterium; life cycle; activator;  
KW bacteria; inhibitor; duodenal ulcer disease; chronic gastritis;  
KW diagnosis.  
OS Helicobacter pylori.  
FH Key Location/Qualifiers  
FT misc\_difference 6 /label= unknown  
FT /note= "encoded by ARC"  
PN W09640893-A1.  
PD 19-DEC-1996.  
PF 06-JUN-1996; U09122.  
PR 07-JUN-1995; US-487032.  
PR 01-APR-1996; US-630405.  
PA (ASTR ) ASTRA AB.  
PI Berglindh OT, Smith D, Mellgaard BL;  
DR WPI; 97-052306/05.  
DR N-PSDB; TG7824.  
PT Helicobacter pylori nucleic acid sequences and related  
PT polypeptide(s) - useful for vaccines to treat or prevent H. pylori  
PT infection, and to detect Helicobacter  
PS Claim 56; Page 707-708; 1481pp; English.  
CC This sequence shows a Helicobacter pylori cell envelope protein  
CC that may be used in a vaccine to prevent or treat H. pylori  
CC infection or to identify H. pylori polypeptide binding compounds,  
CC useful as potential H. pylori life cycle activators or inhibitors.

FT Misc\_difference /note= "encoded by CCN"  
FT 752 /note= "encoded by CCN"  
FT Misc\_difference /note= "encoded by CCN"  
FT 755 /note= "encoded by CCN"  
FT Misc\_difference /note= "encoded by CCN"  
FT 758 /note= "encoded by CCN"  
FT Misc\_difference /note= "encoded by CCN"  
FT 771 /note= "encoded by CCN"  
FT Misc\_difference /note= "encoded by CCN"  
FT 775 /note= "encoded by CCN"  
FT Misc\_difference /note= "encoded by CCN"  
FT 781 /note= "encoded by CCN"  
FT Misc\_difference /note= "encoded by CCN"  
FT 835 /note= "encoded by CCN"  
FT Misc\_difference /note= "encoded by CCN"  
FT 843 /note= "encoded by CCN"  
FT Misc\_difference /note= "encoded by CCN"  
FT 847 /note= "encoded by CCN"  
FT Misc\_difference /note= "encoded by CCN"  
FT 849 /note= "encoded by CCN"  
FT Misc\_difference /note= "encoded by CCN"  
FT 850 /note= "encoded by CCN"  
FT Misc\_difference /note= "encoded by CCN"  
FT 868 /note= "encoded by CCN"  
FT Misc\_difference /note= "encoded by CCN"  
FT 872 /note= "encoded by CCN"  
FT Misc\_difference /note= "encoded by CCN"  
FT 874 /note= "encoded by CCN"  
FT Misc\_difference /note= "encoded by CCN"  
FT 876 /note= "encoded by CCN"  
FT Misc\_difference /note= "encoded by CCN"  
FT 894 /note= "encoded by CCN"  
FT Misc\_difference /note= "encoded by CCN"  
FT 897 /note= "encoded by CCN"  
FT Misc\_difference /note= "encoded by CCN"  
FT 898 /note= "encoded by CCN"  
FT Misc\_difference /note= "encoded by CCN"  
FT 904 /note= "encoded by CCN"  
FT Misc\_difference /note= "encoded by CCN"  
FT 917 /note= "encoded by CCN"  
FT Misc\_difference /note= "encoded by CCN"  
FT 918 /note= "encoded by CCN"  
FT Misc\_difference /note= "encoded by CCN"  
FT 925 /note= "encoded by CCN"  
FT Misc\_difference /note= "encoded by CCN"  
FT 926 /note= "encoded by CCN"  
FT Misc\_difference /note= "encoded by CCN"  
FT 934 /note= "encoded by CCN"  
FT Misc\_difference /note= "encoded by CCN"  
FT 940 /note= "encoded by CCN"  
FT Misc\_difference /note= "encoded by CCN"  
FT 942 /note= "encoded by CCN"  
FT Misc\_difference /note= "encoded by CCN"  
FT 943 /note= "encoded by CCN"  
FT Misc\_difference /note= "encoded by CCN"  
FT 950 /note= "encoded by CCN"  
FT Misc\_difference /note= "encoded by CCN"  
FT 954 /note= "encoded by CCN"  
FT Misc\_difference /note= "encoded by CCN"  
FT 955 /note= "encoded by CCN"  
FT Misc\_difference /note= "encoded by CCN"  
FT 960 /note= "encoded by CCN"  
FT Misc\_difference /note= "encoded by CCN"  
FT 963 /note= "encoded by CCN"  
FT Misc\_difference /note= "encoded by CCN"  
FT 974 /note= "encoded by CCN"  
FT Misc\_difference /note= "encoded by CCN"  
FT 978 /note= "encoded by CCN"  
FT Misc\_difference /note= "encoded by CCN"  
FT 993 /note= "encoded by CCN"  
FT Misc\_difference /note= "encoded by CCN"  
FT 994 /note= "encoded by CCN"  
FT ..

Misc\_difference 997 /note= "encoded by CCN"  
FT Misc\_difference 1003 /note= "encoded by CCN"  
FT Misc\_difference 1004 /note= "encoded by CCN"  
FT Misc\_difference 1005 /note= "encoded by CCN"  
FT Misc\_difference 1018 /note= "encoded by CCN"  
FT Misc\_difference 1021 /note= "encoded by CCN"  
FT Misc\_difference 1031 /note= "encoded by CCN"  
FT Misc\_difference 1050 /note= "encoded by CCN"  
FT Misc\_difference 1052 /note= "encoded by CCN"  
FT Misc\_difference 1057 /note= "encoded by CCN"  
FT Misc\_difference 1061 /note= "encoded by CCN"  
FT Misc\_difference 1065 /note= "encoded by CCN"  
FT Misc\_difference 1067 /note= "encoded by CCN"  
FT Misc\_difference 1080 /note= "encoded by CCN"  
FT Misc\_difference 1082 /note= "encoded by CCN"  
FT Misc\_difference 1097 /note= "encoded by CCN"  
FT ..  
Note: remainder of annotations omitted.

Query Match 70.8%; Score 51; DB 1; Length 1311;  
Best Local Similarity 55.6%; Pred. No. 2.03e+02;  
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 1145 FRFNYSNF 1153  
|:|:|:|  
QY 1 FPFYAAAF 9

RESULT 4  
ID Y12022 standard; Protein; 62 AA.  
AC Y12022;  
DT 18-JUN-1999 (first entry)  
DE Human 5; EST secreted protein SEQ ID NO: 335.  
KW Human; secreted protein; EST; expressed sequence tag; diagnosis;  
KW forensic; gene therapy; chromosome mapping; signal peptide;  
KW upstream regulatory sequence; cytokine activity; cell proliferation;  
KW differentiation; haematopoiesis regulation; tissue growth regulation;  
KW reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;  
KW thrombolytic; anti-inflammatory; tumour inhibition.  
OS Homo sapiens.  
PN WO906554-A2.  
PD 11-FEB-1999.  
PF 31-JUL-1998; IB1238.  
PR 01-AUG-1997; US-905134.  
PA (GEST ) GENSET.  
PI Duclert A, Dumas Milne Edwards J, Lacroix B;  
DR WPI; 99-153784/13.  
DR N-PSDB; X40855.  
PT New nucleic acids encoding human secreted proteins - obtained from  
PT cDNA libraries prepared from kidney, fetal kidney, dystrophic  
PT muscle, muscle and heart tissue  
PS Claim 34; Page 473-474; 622pp; English.  
CC X40826 to X41093 represent 5' expressed sequence tags (ESTs) for human  
CC secreted proteins, and encode the proteins given in Y01602 and  
CC Y11994 to Y12260, respectively. The proteins given represent the signal  
CC peptide and an N-terminal fragment of a secreted protein. The nucleic  
CC acid sequences can be used for producing secreted human gene products.





\*\*\*\*\*  
 M P S R L  
 (TM)  
 \*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
 Copyright (c) 1993-1998 University of Edinburgh, U.K.  
 Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm  
 Run on: Fri Apr 14 22:59:42 2000; MasPar time 5.21 Seconds  
 Tabular output not generated.  
 40.906 Million cell updates/sec.

Title: >US-08-452-843-1  
 Description: (1-9) from US08452843.pap  
 Perfect Score: 72  
 Sequence: 1 FPFKYAAAF 9

Scoring table: PAM 150  
 Gap 15

Searched: 188963 seqs, 23686106 residues

Post-processing: Minimum Match 0%  
 Listing first 45 summaries

Database: a:geneseq36  
 1:geneseqp

Statistics: Mean 16.674; Variance 60.979; scale 0.273

Pred. No. is the number of results predicted by chance to have a  
 score greater than or equal to the score of the result being printed,  
 and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description	Pred. No.
1	72	100.0	9	R89362	Immunogenic peptide, b	2.05e+00
2	51	70.8	185	W71501	Helicobacter polypepti	2.03e+02
3	51	70.8	1311	W79275	Protein kinase GAK.	2.03e+02
4	50	69.4	62	Y12022	Human 5' EST secreted	2.50e+02
5	50	69.4	271	Y05958	E. chafeensis OMP-12 p	2.50e+02
6	50	69.4	293	W86009	Human CAAX processing	2.50e+02
7	50	69.4	326	W20553	H. pylori cell envelop	2.50e+02
8	50	69.4	326	W24686	H. pylori inner membra	2.50e+02
9	50	69.4	329	W86010	Mouse CAAX processing	2.50e+02
10	50	69.4	338	W98105	Guman ras carboxy-term	2.50e+02
11	50	69.4	496	W20801	H. pylori inner membra	2.50e+02
12	50	69.4	511	W75421	T. thermophilus nitrate	2.50e+02
13	50	69.4	589	R32190	Sequence encoded by th	2.50e+02
14	50	69.4	589	R10681	Polyhydroxybutyrate po	2.50e+02
15	50	69.4	672	R17325	Poly-beta-hydroxyalkan	2.50e+02
16	49	68.1	18	R53564	Birch pollen major all	3.08e+02
17	49	68.1	160	R04605	Major Birch allergen B	3.08e+02
18	49	68.1	176	R21796	Bet v I allergen of bi	3.08e+02
19	49	68.1	476	W88461	Human 7-transmembrane	3.08e+02
20	48	66.7	329	W89181	Human RCE1 (hRCE1) pol	3.78e+02
21	48	66.7	655	R89197	Human hepatocellular g	3.78e+02
22	48	66.7	655	R53962	Hepatocyte growth fact	3.78e+02
23	47	65.3	866	W94920	Rat pheromone receptor	4.64e+02

24	47	65.3	1129	1	W50144	Oat phytochrome A apop	4.64e+02
25	47	65.3	4302	1	W00870	Polycystic kidney dise	4.64e+02
26	47	65.3	4302	1	W23396	Human PKD1 polypeptide	4.64e+02
27	47	65.3	4302	1	W23830	Human PKD1 protein.	4.64e+02
28	47	65.3	4303	1	R90302	Polycystic kidney dise	4.64e+02
29	47	65.3	4339	1	R87539	Polycystic kidney dise	4.64e+02
30	47	65.3	4339	1	R75916	Polycystic kidney dise	4.64e+02
31	46	63.9	92	1	W89175	Anti-p53 monoclonal an	5.69e+02
32	46	63.9	97	1	Y11334	S. pneumoniae Arp-depe	5.69e+02
33	46	63.9	274	1	P81275	Human alpha 2-plasmin	5.69e+02
34	46	63.9	274	1	P81275	Human alpha 2-plasmin	5.69e+02
35	46	63.9	464	1	W26426	Swinepox virus HindIII	5.69e+02
36	46	63.9	471	1	R05411	Pro-type human plasma	5.69e+02
37	46	63.9	490	1	P80962	Alpha-2 plasmin inhibi	5.69e+02
38	46	63.9	491	1	R04252	Amino acid sequence of	5.69e+02
39	46	63.9	492	1	R13860	Human alpha-2 plasmin	5.69e+02
40	46	63.9	597	1	P90466	Human alpha-2-plasmin	5.69e+02
41	46	63.9	704	1	Y11682	Sulfated fucose-contai	5.69e+02
42	46	63.9	744	1	P81006	Sulfated fucose-contai	5.69e+02
43	46	63.9	1614	1	R75917	Alpha-2-plasmin inhibi	5.69e+02
44	46	63.9	1614	1	R87538	Polycystic kidney dise	5.69e+02
45	46	63.9	3224	1	W54235	Human Nup358 protein.	5.69e+02

## ALIGNMENTS

RESULT 1  
 ID R89362 standard; peptide; 9 AA.

AC R89362; 1996 (first entry)

DT 18-SEP-1996

DE Immunogenic peptide, based on B35 consensus peptide.

KW Immunogenic peptide; supermotif; HLA molecule; CTL response;

KW therapeutic; diagnostic; cancer; viral infection; hepatitis B;

KW hepatitis C.

OS Synthetic.

PN W09603140-A1.

PD 08-FEB-1996.

PF 21-JUL-1995; U09234.

PR 21-JUL-1994; US-278634.

PR 23-NOV-1994; US-344824.

PR 30-MAY-1995; US-452843.

PA (CYTE-) CYTEL CORP.

PI Sette A, Sidney J.

DR WPI; 96-116784/12.

PT Comps. comprising immunogenic peptide with supermotif allowing more than one HLA mol. to bind - used to induce CTL response in patient and for in vivo and ex vivo therapeutic and diagnostic applications

PS Claim 2; Page 26; 32pp; English.

CC The sequences given in R89362-82 are immunogenic peptides which were use in the composition of the invention. The composition comprises an immunogenic peptide of 9-10 residues with a supermotif which

CC allows binding of more than one HLA molecule. It pref. comprises two conserved residues, a first at the 2nd position from the N-

CC terminal is Pro, and a 2nd at the C-terminal is Met. These peptides are used to induce a CTL response in a patient. They are also

CC useful in compositions for in vivo and ex vivo therapeutic and diagnostic applications, e.g the treatment of cancer and viral

CC infections, e.g. hepatitis B and C.

SQ Sequence 9 AA;

Query Match 100.0%; Score 72; DB 1; Length 9;

Best Local Similarity 100.0%; Pred. NO. 2.05e+00;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 FPFKYAAAF 9

Qy 1 FPFKYAAAF 9

## RESULT 2

ID W71501 standard; Protein; 185 AA.

AC W71501;

DT 09-NOV-1998 (first entry)

THIS PAGE BLANK (uspto)

Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon, E.; Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror, S.J.; Serror, P.; Shin, B.S.; Soldo, B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Takahashi, H.; Takemaru, K.; Takeuchi, M.; Takakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.; Vandenbol, M.; Vannier, F.; Vassarotti, A.; Viari, A.; Wambutt, R.; Wedler, E.; Wedler, H.; Weitzenecker, T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.; Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.

#journal Nature (1997) 390:249-256  
#title The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.

#cross-references MUID:98044033

#accession D70010

##status preliminary; nucleic acid sequence not shown;

##molecule\_type DNA

##residues 1-88 #label KUN

##cross-references GB:Z99120; GB:AL009126; NID:g2635613; PID:el184221;

##experimental\_source strain 168  
PID:g2635639

#### GENETICS

#gene yugE

SUMMARY #length 88 #molecular-weight 10065 #checksum 2766

Query Match 69.4%; Score 50; DB 2; Length 88;

Best Local Similarity 62.5%; Pred. No. 3.23e+01;

Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 21 PFKYGEF 28

QY 2 PFKYAAF 9

#### RESULT 13

ENTRY I65230 #type complete

TITLE DNA-binding protein - human

ORGANISM #formal\_name Homo sapiens #common\_name man

DATE 29-May-1998 #sequence\_revision 29-May-1998 #text\_change

ACCESSIONS I65230

REFERENCE I52306

#authors Luzi, P.; Strayer, D.S.

#journal Biochem. Biophys. Res. Commun. (1995) 208:153-160

#title DNA binding proteins that amplify surfactant protein B gene

expression: isolation and characterization.

#cross-references MUID:95194400

#accession I65230

##status preliminary; translated from GB/EMBL/DBJ

##molecule\_type mRNA

##residues 1-117 #label RES

##cross-references GB:L10405; NID:g860729; PID:g860730

SUMMARY #length 117 #molecular-weight 12849 #checksum 9226

Query Match 69.4%; Score 50; DB 2; Length 117;

Best Local Similarity 62.5%; Pred. No. 3.23e+01;

Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 52 FSKYSAT 59

QY 1 FPKYAAA 8

#### RESULT 14

ENTRY PC2131 #type fragment

TITLE hepatocyte growth factor receptor - rat (fragment)

ORGANISM #formal\_name Rattus norvegicus #common\_name Norway rat

DATE 03-May-1994 #sequence\_revision 07-Oct-1994 #text\_change

ACCESSIONS 17-Mar-1999

PC2131

#### REFERENCE

#authors

PC2131 Tsujii, M.; Kawano, S.; Tsuji, S.; Ito, T.; Hayashi, N.; Horimoto, M.; Mita, E.; Nagano, K.; Masuda, E.; Hayashi, N.; Fusamoto, H.; Kamada, T.

#journal Biochem. Biophys. Res. Commun. (1994) 200:536-541

#title Increased expression of c-met messenger RNA following acute gastric injury in rats.

#cross-references MUID:94220137

#accession PC2131

##molecule\_type mRNA

##residues 1-132 #label TSU

COMMENT This protein participates in the healing process of gastric mucosa after injury.

#### GENETICS

#gene C-met

CLASSIFICATION #superfamily hepatocyte growth factor receptor; protein

kinase homology

KEYWORDS ATP; receptor

SUMMARY #length 132 #checksum 7223

Query Match 69.4%; Score 50; DB 2; Length 132;

Best Local Similarity 55.6%; Pred. No. 3.23e+01;

Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Db 3 FPIKYVND 11

QY 1 FPFKYAAAF 9

#### RESULT 15

ENTRY A38099 #type complete

TITLE glycopeptide N-tetradecanoyltransferase (EC 2.3.1.97) -

Yeast (*Candida albicans*)

ORGANISM #formal\_name Candida albicans

DATE 07-Apr-1994 #sequence\_revision 07-Apr-1994 #text\_change

ACCESSIONS A38099

REFERENCE A38099

#authors Wiegand, R.C.; Carr, C.; Minnerly, J.C.; Pauley, A.M.;

Carron, C.P.; Langner, C.A.; Duronio, R.J.; Gordon, J.I.

#journal J. Biol. Chem. (1992) 267:8591-8598

#title The Candida albicans myristoyl-CoA:protein

N-myristoyltransferase gene. Isolation and expression in

*Saccharomyces cerevisiae* and *Escherichia coli*.

#cross-references MUID:92235090

#accession A38099

##status preliminary; not compared with conceptual translation

##molecule\_type DNA

##residues 1-451 #label WIE

##cross-references GB:M80544; NID:g170883; PID:g170884

KEYWORDS acyltransferase

SUMMARY #length 451 #molecular-weight 51877 #checksum 1063

Query Match 69.4%; Score 50; DB 2; Length 451;

Best Local Similarity 55.6%; Pred. No. 3.23e+01;

Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Db 115 FRKYSHEF 123

QY 1 FPFKYAAAF 9

Search completed: Fri Apr 14 23:00:48 2000

Job time : 10 secs.

```

#submission submitted to the Protein Sequence Database, August 1994
#accession S45923
##molecule_type DNA
##residues 1-404 ##label DOM
##cross-references EMBL:D35932; NID:G536306; PID:G536307; MIPS:YBR063C
##experimental_source strain S288C

GENETICS
#map_position 2R
CLASSIFICATION #superfamily acyl carrier protein homology
KEYWORDS phosphopantetheine; phosphoprotein; transmembrane protein
FEATURE
93-113 #domain transmembrane #status predicted #label TMM\
8 #binding_site phosphopantetheine (Ser) (covalent)
#status predicted
SUMMARY #length 404 #molecular-weight 46444 #checksum 1399

Query Match 70.8%; Score 51; DB 2; Length 404;
Best Local Similarity 44.4%; Pred. No. 2.19e+01;
Matches 4; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 10 FPYEGSDF 18
II: I:: I
QY 1 FPFKYAAF 9

RESULT 10
ENTRY #type complete
TITLE surface-layer glycoprotein precursor - Methanothermus
SOCIALBILIS
ORGANISM #formal_name Methanothermus sociabilis
DATE 22-Jan-1993 #sequence_revision 22-Jan-1993 #text_change
13-Sep-1998
ACCESSIONS S16375; S26144; S26098; S21874
REFERENCE S16225
#authors Broeckl, G.; Behr, M.; Fabry, S.; Hensel, R.; Kaudewitz, H.;
#journal Biendl, E.; Koenig, H.
#title Analysis and nucleotide sequence of the genes encoding the
surface-layer glycoproteins of the hyperthermophilic
methanogens Methanothermus fervidus and Methanothermus
sociabilis.
#cross-references MUID:91293115
#accession S16375
##molecule_type DNA
##residues 1-593 ##label BRO
##cross-references EMBL:X58297; NID:G44281; PID:G809714

GENETICS
#gene slgA
#start_codon GTG
KEYWORDS glycoprotein
FEATURE
1-22 #domain signal sequence #status predicted #label SIG\
23-593 #product surface-layer glycoprotein #status predicted
#label MAT
SUMMARY #length 593 #molecular-weight 65481 #checksum 7421

Query Match 70.8%; Score 51; DB 2; Length 593;
Best Local Similarity 62.5%; Pred. No. 2.19e+01;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 526 YPFKYAVS 533
:||||:
QY 1 FPFKYAAA 8

#cross-references MUID:91293115
#accession S16375
##molecule_type DNA
##residues 1-593 ##label BRO
##cross-references EMBL:X58296
#accession S26144
##molecule_type protein
##residues 23-42 ##label BRO2
REFERENCE S21873
#authors Broeckl, G.
#submission submitted to the EMBL Data Library, March 1991
#accession S26098
##molecule_type DNA
##residues 1-256,'I',258,'V',260-593 ##label BRO1
#cross-references EMBL:X58296; NID:G44546; PID:G809717

GENETICS
#gene slgA
#start_codon GTG
KEYWORDS glycoprotein
FEATURE
1-22 #domain signal sequence #status predicted #label SIG\
23-593 #product surface-layer glycoprotein #status experimental
#label MAT
SUMMARY #length 593 #molecular-weight 65503 #checksum 8058

Query Match 70.8%; Score 51; DB 2; Length 593;
Best Local Similarity 62.5%; Pred. No. 2.19e+01;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 526 YPFKYAVS 533
:||||:
QY 1 FPFKYAAA 8

```

## RESULT 11

```

ENTRY #type complete
TITLE surface-layer glycoprotein precursor - Methanothermus
fervidus
ORGANISM #formal_name Methanothermus fervidus
DATE 22-Jan-1993 #sequence_revision 22-Jan-1993 #text_change
09-Sep-1997
ACCESSIONS S16225; S21873
REFERENCE S16225
#authors Broeckl, G.; Behr, M.; Fabry, S.; Hensel, R.; Kaudewitz, H.;
#journal Biendl, E.; Koenig, H.
#title Analysis and nucleotide sequence of the genes encoding the
surface-layer glycoproteins of the hyperthermophilic
methanogens Methanothermus fervidus and Methanothermus
sociabilis.
#cross-references MUID:91293115
#accession S16225
##molecule_type DNA
##residues 1-593 ##label BRO
##cross-references EMBL:X58297; NID:G44281; PID:G809714

GENETICS
#gene slgA
#start_codon GTG
KEYWORDS glycoprotein
FEATURE
1-22 #domain signal sequence #status predicted #label SIG\
23-593 #product surface-layer glycoprotein #status predicted
#label MAT
SUMMARY #length 593 #molecular-weight 65481 #checksum 7421

Query Match 70.8%; Score 51; DB 2; Length 593;
Best Local Similarity 62.5%; Pred. No. 2.19e+01;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 526 YPFKYAVS 533
:||||:
QY 1 FPFKYAAA 8

#cross-references MUID:91293115
#accession S16225
##molecule_type DNA
##residues 1-593 ##label BRO
##cross-references EMBL:X58297; NID:G44281; PID:G809714

GENETICS
#gene slgA
#start_codon GTG
KEYWORDS glycoprotein
FEATURE
1-22 #domain signal sequence #status predicted #label SIG\
23-593 #product surface-layer glycoprotein #status predicted
#label MAT
SUMMARY #length 593 #molecular-weight 65481 #checksum 7421

Query Match 70.8%; Score 51; DB 2; Length 593;
Best Local Similarity 62.5%; Pred. No. 2.19e+01;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 526 YPFKYAVS 533
:||||:
QY 1 FPFKYAAA 8

```

## RESULT 12

```

ENTRY #type complete
TITLE hypothetical protein yuGE - Bacillus subtilis
ORGANISM #formal_name Bacillus subtilis
DATE 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change
24-Sep-1998
ACCESSIONS D70010
REFERENCE A69580
#authors Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.;
Alloni, G.; Azevedo, V.; Bertero, M.G.; Bessieres, P.;
Bolotin, A.; Borchert, S.; Boriss, R.; Boursier, L.; Brans,
A.; Braun, M.; Brignell, S.C.; Bron, S.; Brouillet, S.;
Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
Choi, S.K.; Codani, J.J.; Conner, I.F.; Cummings, N.J.;
Daniel, R.A.; Denizot, F.; Devine, K.M.; Duesterhoeft, A.;
Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.;
Fabret, C.; Ferrari, E.; Foulger, D.; Fritz, C.; Fujita,
M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galleron, N.; Ghim,
S.Y.; Glaser, P.; Goffeau, A.; Golightly, E.J.; Grandi, G.;
Guiseppi, G.; Guy, B.J.; Haga, K.; Halech, J.; Harwood,
C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.;
Hullo, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karamata, D.;
Kashara, Y.; Klaert-Blanchard, M.; Klein, C.; Kobayashi,
Y.; Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.;
Kurita, K.; Lapidus, A.; Lardinois, S.; Lauber, J.;
Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.;
Mauvel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno,
M.; Moestl, D.; Nakai, S.; Noback, M.; Noone, S.; O'Reilly,
M.; Ogawa, K.; Ogihara, A.; Oudega, B.; Park, S.H.; Parro,
V.; Pohl, T.M.; Portetelle, D.; Porwollik, S.; Prescott,
A.M.; Presecan, E.; Pujic, P.; Purnelle, B.; Rapoport, G.;

```

```

#journal      J. Biol. Chem. (1994) 269:2996-3009
#title        Comparison of myristoyl-CoA:protein N-myristoyltransferases
              from three pathogenic fungi: Cryptococcus neoformans,
              Histoplasma capsulatum, and Candida albicans.
#cross-references MIMD:94132075
#accession    B49993
#status       preliminary
#molecule_type DNA
#residues     1-529 #label LOD
#cross-references GB:L25118; NID:9407694; PID:9407695
GENETICS
#gene         Nmt
#introns      203/2; 464/3
#keywords     acyltransferase
SUMMARY
#length 529 #molecular-weight 59363 #checksum 3672

Query Match      72.2%; Score 52; DB 2; Length 529;
Best Local Similarity 55.6%; Pred. No. 1.48e+01;
Matches          3; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 192 FRNYSAPF 200
|:|:|:|
Qy 1 PPFKYAAAF 9

RESULT 6
ENTRY  A34434 #type complete
TITLE  arylphorin alpha chain precursor - tobacco hornworm
ORGANISM #formal_name Manduca sexta #common_name tobacco hornworm
DATE 15-Jun-1990 #sequence_revision 15-Jun-1990 #text_change
04-Sep-1998
ACCESSIONS A34434
REFERENCE  A34434
#authors  Willott, E.; Wang, X.Y.; Wells, M.A.
#journal  J. Biol. Chem. (1989) 264:19052-19059
#title    cDNA and gene sequence of Manduca sexta arylphorin, an
          aromatic amino acid-rich larval serum protein. Homology to
          arthropod hemocyanins.
#cross-references MIMD:90037032
#accession A34434
#status    preliminary
#molecule_type DNA
#residues  1-702 #label WIL
#cross-references GB:M28396; EMBL:J05092; NID:gi59486; PID:gi59487;
          EMBL:J05093
CLASSIFICATION #superfamily arylphorin
SUMMARY #length 702 #molecular-weight 83866 #checksum 4883

Query Match      72.2%; Score 52; DB 2; Length 702;
Best Local Similarity 44.4%; Pred. No. 1.48e+01;
Matches          3; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 686 PPKFNVPF 594
|:|:|:|
Qy 1 PPFKYAAAF 9

RESULT 7
ENTRY  T00068 #type complete
TITLE  hypothetical protein KIAA0443 - human
ORGANISM #formal_name Homo sapiens #common_name man
DATE 22-Jan-1999 #sequence_revision 22-Jan-1999 #text_change
22-Jan-1999
ACCESSIONS T00068
REFERENCE  214084
#authors  Ishikawa, K.; Nagase, T.; Nakajima, D.; Seki, N.; Ohira, M.;
          Miyajima, N.; Tanaka, A.; Kotani, H.; Nomura, N.; Ohara, O.
#journal  DNA Res. (1997) 4:307-313
#title    Prediction of the coding sequences of unidentified human
          genes. VIII. 78 new cDNA clones from brain which code for
          large proteins in vitro.
#accession T00068
#status    preliminary; translated from GB/EMBL/DBJ

```

```

#molecule_type mRNA
#residues     1-1395 #label ISH
#cross-references EMBL:AB007903; NID:d1175359; PID:d1024620
#experimental_source brain; clone HJ0137
GENETICS
#note        KIAA0443
SUMMARY #length 1395 #molecular-weight 156836 #checksum 6541

Query Match      72.2%; Score 52; DB 2; Length 1395;
Best Local Similarity 44.4%; Pred. No. 1.48e+01;
Matches          4; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 1111 FPFQYDPSY 1119
|:|:|:|
Qy 1 PPFKYAAAF 9

RESULT 8
ENTRY  E64710 #type complete
TITLE  hypothetical protein HPI1525 - Helicobacter pylori (strain
26695)
ORGANISM #formal_name Helicobacter pylori
DATE 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change
18-Sep-1998
ACCESSIONS E64710
REFERENCE  A64520
#authors  Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.;
          Sutton, G.G.; Fleischmann, R.D.; Ketchum, K.A.; Klenk,
          H.P.; Gill, S.; Dougherty, B.A.; Nelson, K.; Quackenbush,
          J.; Zhou, L.; Kirkness, E.F.; Peterson, S.; Loftus, B.;
          Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.;
          McKenney, E.K.; Fitzgerald, L.M.; Lee, N.; Adams, M.D.;
          Hickey, E.K.; Berg, D.E.; Gocayne, J.D.; Utterback, T.R.;
          Peterson, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.;
          Fujii, C.; Bowman, C.; Watthey, L.; Wallin, E.; Hayes,
          W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser,
          C.M.; Venter, J.C.
#journal  Nature (1997) 388:539-547
#title    The complete genome sequence of the gastric pathogen
          Helicobacter pylori.
#cross-references MIMD:97394467
#accession E64710
#status    preliminary; nucleic acid sequence not shown;
          translation not shown
#molecule_type DNA
#residues  1-211 #label TOM
#cross-references GB:AE000650; GB:AE000511; NID:g2314700; PID:g2314707;
          TIGR:HP1525
GENETICS
#start_codon GTG
CLASSIFICATION #superfamily hypothetical protein HPI066
SUMMARY #length 211 #molecular-weight 24866 #checksum 4606

Query Match      70.8%; Score 51; DB 2; Length 211;
Best Local Similarity 55.6%; Pred. No. 2.19e+01;
Matches          5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 189 FAREYDSAF 197
|:|:|:|
Qy 1 PPFKYAAAF 9

RESULT 9
ENTRY  S45923 #type complete
TITLE  probable phosphatidylserine-binding protein - yeast
          (Saccharomyces cerevisiae)
ALTERNATE_NAMES hypothetical protein YBR0610; hypothetical protein YBR063c
ORGANISM #formal_name Saccharomyces cerevisiae
DATE 26-Aug-1994 #sequence_revision 09-Sep-1994 #text_change
05-Dec-1998
ACCESSIONS S45923
REFERENCE  S45616
#authors  Domdey, H.; Gassenhuber, H.; Obermaier, B.; Piravandi, E.

```

```

ENTRY      S74825      #type complete
TITLE      hypothetical protein alr1747 - Synechocystis sp. (strain PCC
            6803)
ORGANISM   #formal_name Synechocystis sp.
#variety
DATE       PCC 6803
25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change
S74825
S74322
#authors   Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.;
            Nakamura, Y.; Miyajima, N.; Hirose, M.; Sugita, M.;
            Sasamoto, S.; Kimura, T.; Hosouchi, T.; Matsuno, A.;
            Muraki, A.; Nakazaki, N.; Naruo, K.; Okumura, S.; Shimo,
            S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.;
            Yasuda, M.; Tabata, S.
#journal   DNA Res. (1996) 3:109-136
#title     Sequence analysis of the genome of the unicellular
            cyanobacterium Synechocystis sp. PCC6803. II. Sequence
            determination of the entire genome and assignment of
            potential protein-coding regions.
#cross-references MUID:97061201
#accession S74825
#status    nucleic acid sequence not shown; translation not shown
#molecule_type DNA
#residues  1-469 #label KAN
#cross-references EMBL:D90909; GB:AB001339; NID:G1652844; PID:dl018519;
            PID:G1652868
#note      the nucleotide sequence was submitted to the EMBL Data
            Library, June 1996
SUMMARY    #length 469 #molecular-weight 52543 #checksum 1395
Query Match 76.4%; Score 55; DB 2; Length 469;
Best Local Similarity 55.6%; Pred. No. 4.43e+00;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
Db 282 FPFKPSKF 290
QY 1 FPFKYAAAF 9
|||||:|
|||||:|

ENTRY      3
TITLE      #type complete
            alpha-toxin - Clostridium novyi (ATCC 19402)
ORGANISM   #formal_name Clostridium novyi
#variety   ATCC 19402
DATE       28-Oct-1996 #sequence_revision 08-Nov-1996 #text_change
ACCESSIONS S55805; S71294; S71158; S44273; I40834; S44272
REFERENCE   Hofmann, F.; Herrmann, A.; Habermann, E.; von
            Eichel-Streiber, C.
            Mol. Gen. Genet. (1995) 247:670-679
#journal   Sequencing and analysis of the gene encoding the alpha-toxin
            of Clostridium novyi proves its homology to toxins A and B
            of Clostridium difficile.
#cross-references MUID:95342160
#accession S55805
#status    nucleic acid sequence not shown
#molecule_type DNA
#residues  1-2178 #label HOF
#cross-references EMBL:Z48636; NID:G728537; PID:G755724
#accession S71294
#molecule_type protein
#residues  1-15 #label HOW
REFERENCE   S71158
#authors   Hofmann, F.
#submission submitted to the EMBL Data Library, March 1995
#accession S71158
#molecule_type DNA
#residues  1-1179, 'LKV', '1183', 'LVTHIGE', '1191-2178 #label HOS
#cross-references EMBL:Z48636; NID:G728537; PID:G755724
REFERENCE   S44272
#authors   Hofmann, F.; Habermann, E.; von Eichel-Streiber, C.

```

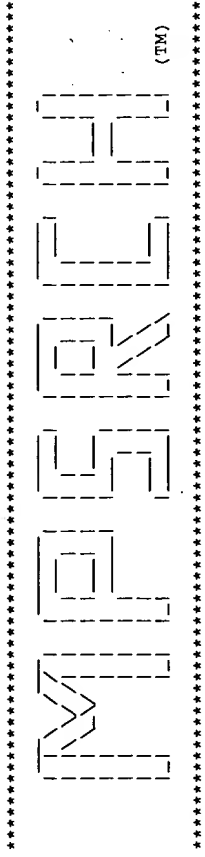
```

#submission submitted to the EMBL Data Library, July 1993
#description Sequence analysis of Clostridium novyi alpha-toxin: a member
            of the family of large clostridial cytotoxins.
#accession S44273
#molecule_type DNA
#residues  1-243;1204-2178 #label HOA
#cross-references EMBL:Z23281
GENETICS   tcn-alpha
#gene       #superfamily cpl repeat homology
KEYWORDS   virulence factor
FEATURE    #domain cpl repeat homology #label cpl2
1880-1899  #length 2178 #molecular-weight 250166 #checksum 5975
SUMMARY
Query Match 75.0%; Score 54; DB 2; Length 2178;
Best Local Similarity 66.7%; Pred. No. 6.65e+00;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
Db 1273 FPMKYEAPF 1281
QY 1 FPFKYAAAF 9
|||||:|
|||||:|

RESULT     4
ENTRY      S73843      #type complete
TITLE      general amino acid permease GAP1 homolog F10_orf503 -
            Mycoplasma pneumoniae (ATCC 29342) (SGC3)
ALTERNATE_NAMES
ORGANISM   #formal_name Mycoplasma pneumoniae
#variety   ATCC 29342
DATE       27-Feb-1997 #sequence_revision 25-Apr-1997 #text_change
ACCESSIONS S73843
REFERENCE   S73327
#authors   Himmelreich, R.; Hilbert, H.; Plagens, H.; Pirk, E.; Li,
            B.C.; Herrmann, R.
#journal   Nucleic Acids Res. (1996) 24:4420-4449
#title     Complete sequence analysis of the genome of the bacterium
            Mycoplasma pneumoniae.
#cross-references MUID:97105885
#accession S73843
#status    preliminary; nucleic acid sequence not shown;
            translation not shown
#molecule_type DNA
#residues  1-503 #label HIM
#cross-references EMBL:AF000051; GB:U00089; NID:G1674211; PID:G1674212
#note      the nucleotide sequence was submitted to the EMBL Data
            Library, November 1996
GENETICS   #gene       gap1
#genetic_code SGC3
SUMMARY    #length 503 #molecular-weight 54960 #checksum 5162
Query Match 72.2%; Score 52; DB 2; Length 503;
Best Local Similarity 55.6%; Pred. No. 1.48e+01;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
Db 23 FAFNVVAGF 31
QY 1 FPFKYAAAF 9
|||||:|
|||||:|

RESULT     5
ENTRY      B49993      #type complete
TITLE      glycylpeptide N-tetradecanoyltransferase (EC 2.3.1.97) -
            Ajellomyces capsulata
ORGANISM   #formal_name Ajellomyces capsulata
DATE       10-Nov-1995 #sequence_revision 10-Nov-1995 #text_change
09-Sep-1997
ACCESSIONS B49993
REFERENCE   B49993
#authors   Lodge, J.K.; Johnson, R.L.; Weinberg, R.A.; Gordon, J.I.

```



Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Apr 14 23:00:38 2000; Maspar time 3.25 Seconds  
111.119 Million cell updates/sec

Tabular output not generated.

Title: >US-08-452-843-1  
Description: (1-9) from US08452843.pap  
Perfect Score: 72  
Sequence: 1 PPFKYAAAF 9

Scoring table: PAM 150  
Gap 15

Searched: 122810 seqs, 40068593 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: p1r62  
1.p1r1 2.p1r2 3.p1r3 4.p1r4

Statistics: Mean 23.702; Variance 35.510; scale 0.667

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	55	76.4	173	2 B69143	hypothetical protein	4.43e+00
2	55	76.4	469	2 S74825	hypothetical protein	4.43e+00
3	54	75.0	2178	2 S58005	alpha-toxin - Clostri	6.65e+00
4	52	72.2	503	2 S73843	general amino acid pe	1.48e+01
5	52	72.2	529	2 B49993	glycylpeptide N-tetra	1.48e+01
6	52	72.2	702	2 A34434	arylphorin alpha chai	1.48e+01
7	52	72.2	1395	2 T00068	hypothetical protein	1.48e+01
8	51	70.8	211	2 E64710	hypothetical protein	2.19e+01
9	51	70.8	404	2 S45923	probable phosphopante	2.19e+01
10	51	70.8	593	2 S16375	surface-layer glycopr	2.19e+01
11	51	70.8	593	2 S16225	surface-layer glycopr	2.19e+01
12	50	69.4	88	2 D70010	hypothetical protein	3.23e+01
13	50	69.4	117	2 I65230	DNA-binding protein -	3.23e+01
14	50	69.4	132	2 PC2131	hepatocyte growth fac	3.23e+01
15	50	69.4	451	2 A38099	glycylpeptide N-tetra	3.23e+01
16	50	69.4	453	2 B70426	periplasmic serine pr	3.23e+01
17	50	69.4	490	2 A64679	NADH dehydrogenase (u	3.23e+01
18	50	69.4	492	2 G71839	nadh oxidoreductase I	3.23e+01
19	50	69.4	549	2 B65215	hypothetical 59.2 kD	3.23e+01
20	50	69.4	589	2 A34341	poly(3-hydroxybutyrat	3.23e+01
21	50	69.4	664	2 S73624	hypothetical protein	3.23e+01
22	50	69.4	840	2 A42970	H+-transporting ATPas	3.23e+01
23	50	69.4	890	2 S54554	H+-transporting ATPas	3.23e+01

24	50	69.4	1004	2 A48821	Wnt-5 protein - fruit	3.23e+01
25	50	69.4	1379	2 S01354	hepatocyte growth fac	3.23e+01
26	50	69.4	1390	1 TVHUME	hepatocyte growth fac	3.23e+01
27	49	68.1	160	2 F55699	major pollen allergen	4.73e+01
28	49	68.1	160	2 C55699	major pollen allergen	4.73e+01
29	49	68.1	160	2 T55699	major pollen allergen	4.73e+01
30	49	68.1	160	2 S05376	major pollen allergen	4.73e+01
31	49	68.1	160	2 E55699	major pollen allergen	4.73e+01
32	49	68.1	160	2 G55699	major pollen allergen	4.73e+01
33	49	68.1	291	2 H69521	4-hydroxybenzoate oct	4.73e+01
34	49	68.1	355	2 A32115	stearyl-CoA desaturase	4.73e+01
35	49	68.1	391	2 S72717	Leppb1170.F3_112 prote	4.73e+01
36	49	68.1	427	2 S73659	MG288 homolog P02.Orf	4.73e+01
37	49	68.1	545	2 B44054	orf2 protein - Junoni	4.73e+01
38	49	68.1	607	2 C40361	virC-region hypothereti	4.73e+01
39	49	68.1	748	2 A45243	envelope protein HrpH	4.73e+01
40	49	68.1	822	2 H69547	molybdopterin oxidore	4.73e+01
41	49	68.1	1342	2 A31946	xanthine dehydrogenase	4.73e+01
42	49	68.1	2470	2 I50726	cation-independent ma	4.73e+01
43	48	66.7	211	2 B71809	hypothetical protein	6.89e+01
44	48	66.7	276	2 A70425	hypothetical protein	6.89e+01
45	48	66.7	585	2 S77114	ABC1-type transport p	6.89e+01

ALIGNMENTS

RESULT 1

ENTRY B69143 #type complete

TITLE hypothetical protein MTH336 - Methanobacterium thermoautotrophicum (strain Delta H)

ORGANISM #formal\_name Methanobacterium thermoautotrophicum

DATE 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 05-Feb-1999

ACCESSIONS B69143

REFERENCE A69000

#authors Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.; Bashirzadeh, R.; Blakely, D.; Cook, R.; Gilbert, K.; Harrison, D.; Hoang, L.; Keagle, P.; Lumm, W.; Pothier, B.; Qiu, D.; Spadafora, R.; Vicalre, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwan, N.; Caruso, A.; Bush, D.; Safer, H.; Patwell, D.; Prabhakar, S.; Choudhary, S.; Shimer, G.; Goyal, A.; Petrokovski, S.; McBratney, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.

#journal J. Bacteriol. (1997) 179:7135-7155

#title Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: functional analysis and comparative genomics.

#cross-references MUID:98037514

#accession B69143

#status preliminary: nucleic acid sequence not shown; translation not shown

#molecule\_type DNA

#residues 1-173 #label MTH

#cross-references GB:A500818; GB:AE000666; NID:g2621384; PID:g2621392

#experimental\_source strain Delta H

GENETICS

#gene MTH336

#start\_codon TTG

CLASSIFICATION #superfamily Methanococcus jannaschii conserved hypothetical protein MJ207

SUMMARY #length 173 #molecular-weight 20426 #checksum 9042

Query Match 76.4%; Score 55; DB 2; Length 173;  
Best Local Similarity 55.68; Pred. No. 4.43e+00;  
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 32 PPFKYPLVF 40  
|||:|  
QY 1 PPFKYAAAF 9

RESULT 2



**THIS PAGE BLANK (USPTO)**

DR PIR: JQ0757; JQ0757.  
 DR PROSITE; PS00322; HISTONE\_H3\_1; 1.  
 DR PROSITE; PS00959; HISTONE\_H3\_2; 1.  
 DR PFAM; PF00125; histone; 1.  
 KW Nuclear protein; Chromosomal protein; DNA-binding; Nucleosome core.  
 FT INIT\_MET 0  
 SQ SEQUENCE 135 AA; 15154 MW; 99B89C5B CRC32;

Query Match 100.0%; Score 69; DB 1; Length 135;  
 Best Local Similarity 100.0%; Pred. No. 2.46e-05;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 40 RYRPGTVAL 48  
 QY 1 RYRPGTVAL 9

Search completed: Sat Apr 15 00:02:33 2000  
 Job time : 45 secs.

RP SEQUENCE FROM N.A. (F55G1.2).  
RC STRAIN-BRISTOL N2;  
RA WATERSTON R.;  
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.  
RN (4)  
RP SEQUENCE FROM N.A. (B0035.10).  
RC STRAIN-BRISTOL N2;  
RA WHITE S.;  
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.  
RN (5)  
RP SEQUENCE FROM N.A. (K06C4.5 AND K06C4.13).  
RC STRAIN-BRISTOL N2;  
RA MILLER N., BRADSHAW H.;  
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.  
RN (6)  
RP SEQUENCE FROM N.A. (F45F2.13).  
RC STRAIN-BRISTOL N2;  
RA DAVIDSON S., WOHLDMANN P.;  
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.  
RN (7)  
RP SEQUENCE FROM N.A. (F54E12.1).  
RC STRAIN-BRISTOL N2;  
RA WHITE S., MORTIMORE B.;  
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.  
RN (8)  
RP SEQUENCE FROM N.A. (ZK131.2; ZK131.3 AND ZK131.7).  
RC STRAIN-BRISTOL N2;  
RA STEWARD C.;  
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.  
RN (9)  
RP SEQUENCE.  
RC STRAIN-DR27;  
RX MEDLINE; 87105951.  
RA VANFLTEREN J.R., VAN BUN S.M., VAN BEEUMEN J.J.;  
RT "The primary structure of histone H3 from the nematode *Caenorhabditis elegans*," 211:59-63(1987).  
RL FEBS Lett. 211:59-63(1987).  
CC -1- FUNCTION: HISTONE H3, ALONG WITH HISTONE H4, PLAYS A CENTRAL ROLE  
CC -1- IN NUCLEOSOME FORMATION.  
CC -1- SUBUNIT: THE NUCLEOSOME IS AN OCTAMER CONTAINING TWO MOLECULES OF  
CC H2A, H2B, H3, AND H4; WHICH WRAP APPROXIMATIVELY 146 BP OF DNA.  
CC -1- SIMILARITY: BELONGS TO THE HISTONE H3 FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; X15634; CAA33644.1; -  
CC EMBL; 268336; CAA92733.1; -  
CC EMBL; U58750; AAB00650.1; -  
CC EMBL; U64843; AAB04857.1; -  
CC EMBL; U64843; AAB04852.1; -  
CC EMBL; U64845; AAC48033.1; -  
CC EMBL; 282271; CAB05209.1; -  
CC EMBL; 283245; CAB05831.1; -  
CC EMBL; 283245; CAB05833.1; -  
CC EMBL; 283245; CAB05834.1; -  
CC EMBL; 273102; CAA97411.1; -  
CC PIR; A25842; HSKW3.  
CC PIR; S04241; S04241.  
CC WORMPEP; B0035.10; CE03253.  
CC WORMPEP; F22B3.2; CE03253.  
CC WORMPEP; F45F2.13; CE10540.  
CC WORMPEP; F55G1.2; CE10540.  
CC WORMPEP; F55G1.2; CE03253.  
CC WORMPEP; K06C4.5; CE10540.  
CC WORMPEP; K06C4.13; CE10540.  
CC WORMPEP; ZK131.2; CE10540.  
CC WORMPEP; ZK131.3; CE10540.  
CC -----  
DR WORMPEP; ZK131.7; CE10540.  
DR PROSITE; PS00322; HISTONE\_H3\_1; 1.  
DR PROSITE; PS00959; HISTONE\_H3\_2; 1.  
DR PRAM; PF00125; histone; 1.  
KW Nuclear protein; Chromosomal protein; DNA-binding; Nucleosome core;  
FT INIT\_MET 0 0  
FT MOD\_RES 4 4 ACETYLATION.  
FT MOD\_RES 14 14 ACETYLATION.  
FT MOD\_RES 23 23 METHYLATION.  
FT MOD\_RES 9 9 METHYLATION.  
FT MOD\_RES 27 27 METHYLATION.  
FT MOD\_RES 36 36 METHYLATION.  
FT MOD\_RES 79 79 METHYLATION.  
FT MOD\_RES 96 96 CARBOXYMETHYLATION.  
FT VARIANT 96 96 A -> C.  
FT VARIANT 100 100 L -> I.  
FT CONFLICT 64 64 R -> K (IN REF. 3).  
SQ SEQUENCE 135 AA; 15213 MW; 43A8A478 CRC32;  
Query Match 100.0%; Score 69; DB 1; Length 135;  
Best Local Similarity 100.0%; Pred. No. 2.46e-05;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Db 40 RYRPGTVAL 48  
QY 1 RYRPGTVAL 9  
|||||  
RESULT 15  
ID H3\_ACRFO STANDARD; PRT; 135 AA.  
AC P22843;  
DT 01-AUG-1991 (Rel. 19, Created)  
DT 01-AUG-1991 (Rel. 19, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE HISTONE H3.  
OS Acropora formosa (Staghorn coral).  
OC Eukaryota; Metazoa; Chordata; Anthozoa; Zoantharia; Scleractinia;  
OC Acroporidae; Acropora.  
RN [1]  
RX SEQUENCE FROM N.A.  
RX MEDLINE; 91033046.  
RA MILLER D.J., MCILLAN J., MILES A., TEN LOHUIS M., MAHONY T.;  
RT "Nucleotide sequence of the histone H3-encoding gene from the  
RT scleractinian coral *Acropora formosa* (Cnidaria: Scleractinia).";  
RL Gene 93:319-320(1990).  
RN [2]  
RX SEQUENCE FROM N.A.  
RX MEDLINE; 94047119.  
RA MILLER D.J., HARRISON P.L., MAHONY T.J., MCILLAN J.P., MILES A.,  
RA ODORICO D.M., TEN LOHUIS M.R.;  
RT "Nucleotide sequence of the histone gene cluster in the coral  
RT *Acropora formosa* (Cnidaria: Scleractinia): features of histone gene  
RT structure and organization are common to diploblastic and  
RT triploblastic metazoans.";  
RL J. Mol. Evol. 37:245-253(1993).  
CC -1- FUNCTION: HISTONE H3, ALONG WITH HISTONE H4, PLAYS A CENTRAL ROLE  
CC IN NUCLEOSOME FORMATION.  
CC -1- SUBUNIT: THE NUCLEOSOME IS AN OCTAMER CONTAINING TWO MOLECULES OF  
CC H2A, H2B, H3, AND H4; WHICH WRAP APPROXIMATIVELY 146 BP OF DNA.  
CC -1- SIMILARITY: BELONGS TO THE HISTONE H3 FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; M60509; AAA64958.1; -  
CC EMBL; L11067; AAC37352.1; -  
CC EMBL; S67324; AAB28736.1; -

RT H3 histone-encoding gene family.";  
RL Nucleic Acids Res. 19:6327-6327(1991).  
RN [11]  
RP SEQUENCE.  
RC SPECIES-CHICKEN;  
RX MEDLINE; 74308333.  
RA BRANDT W.F., VON HOLT C.;  
RT "The determination of the primary structure of histone F3 from  
RT chicken erythrocytes by automatic Edman degradation. 2. Sequence  
RT analysis of histone F3.";  
RL Eur. J. Biochem. 46:419-429(1974).  
RN [12]  
RP SEQUENCE FROM N.A.  
RC SPECIES-C.MOSCHATA;  
RX MEDLINE; 89178747.  
RA TOENIES R., MUNK K., DOENECKE D.;  
RT "Conserved organization of an avian histone gene cluster with  
RT inverted duplications of H3 and H4 genes.";  
RL J. Mol. Evol. 28:200-211(1989).  
RN [13]  
RP SEQUENCE.  
RC SPECIES-I.BUBALUS;  
RX MEDLINE; 73165575.  
RA HOOPER J.A., SMITH E.L., SOMMER K.R., CHALKLEY R.;  
RT "Histone 3. IV. Amino acid sequence of histone 3 of the testes of the  
RT carp, *Letiobus bubalus*.";  
RL J. Biol. Chem. 248:3275-3279(1973).  
RN [14]  
RP SEQUENCE.  
RC SPECIES-P.AFRICANUS;  
RX MEDLINE; 74309063.  
RA BRANDT W.F., STRICKLAND W.N., VON HOLT C.;  
RT "The primary structure of histone F3 from shark erythrocytes.";  
RL FEBS Lett. 40:349-352(1974).  
RN [15]  
RP SEQUENCE FROM N.A.  
RC SPECIES-O.MYKISS;  
RX MEDLINE; 85083109.  
RA CONNOR W., STATES J.C., MEZQUITA J., DIXON G.H.;  
RT "Organization and nucleotide sequence of rainbow trout histone H2A  
RT and H3 genes.";  
RL J. Mol. Evol. 20:236-250(1984).  
RN [16]  
RP SEQUENCE OF 1-25.  
RC SPECIES-O.MYKISS;  
RX MEDLINE; 72259090.  
RA CANDIDO E.P.M., DIXON G.H.;  
RT "Amino-terminal sequences and sites of in vivo acetylation of trout-  
RT testis histones 3 and IIB 2.";  
RL Proc. Natl. Acad. Sci. U.S.A. 69:2015-2019(1972).  
RN [17]  
RP SEQUENCE FROM N.A. (CLONE XLHW23).  
RC SPECIES-X.LAEVIS;  
RX MEDLINE; 86041919.  
RA OLD R.W., SHEIKH S.A., CHAMBERS A., NEWTON C.A., MOHAMMED A.,  
RT ALDRIDGE T.C.;  
RT "Individual Xenopus histone genes are replication-independent in  
RT oocytes and replication-dependent in Xenopus or mouse somatic  
RT cells.";  
RL Nucleic Acids Res. 13:7341-7358(1985).  
RN [18]  
RP SEQUENCE FROM N.A. (GENE CLUSTERS X1H1 AND X1H3).  
RC SPECIES-X.LAEVIS;  
RX MEDLINE; 86037224.  
RA PERRY M., THOMSEN G.H., ROEDER R.G.;  
RT "Genomic organization and nucleotide sequence of two distinct histone  
RT gene clusters from *Xenopus laevis*. Identification of novel conserved  
RT upstream sequence elements.";  
RL J. Mol. Biol. 185:479-499(1985).  
RN [19]  
RP SEQUENCE FROM N.A.  
RC SPECIES-P.DUMERILLII; TISSUE-SPERM;  
RX MEDLINE; 90306006.

RA SELLOS D., KRAWETZ S.A., DIXON G.H.;  
RT "Organization and complete nucleotide sequence of the  
RT core-histone-gene cluster of the annelid *Platynereis dumerillii*.";  
RL Eur. J. Biochem. 190:21-29(1990).  
RN [20]  
RP SEQUENCE FROM N.A.  
RC SPECIES-D.HYDEI;  
RX MEDLINE; 90221886.  
RA KREMER H., HENNIG W.;  
RT "Isolation and characterization of a *Drosophila hydei* histone DNA  
RT repeat unit.";  
RL Nucleic Acids Res. 18:1573-1580(1990).  
RN [21]  
RP SEQUENCE FROM N.A.  
RC SPECIES-D.HYDEI;  
RA STRAUSBAUGH L.D., FITCH D.H.A., BARRETT V.;  
RL Submitted (APR-1990) to the EMBL/GenBank/DBJ databases.  
RN [22]  
RP SEQUENCE FROM N.A.  
RC SPECIES-T.CALIFORNICUS;  
RX MEDLINE; 92127060.  
RA PORTER D., BROWN D., WELLS D.;  
RT "An H3-H4 histone gene pair in the marine copepod *Tigriopus*  
RT californicus, contains an intergenic dyad symmetry element.";  
RL DNA Seq. 1:197-206(1991).  
RN [23]  
RP SEQUENCE FROM N.A.  
RC SPECIES-T.CALIFORNICUS;  
RX MEDLINE; 93076000.  
RA BROWN D., COOK A., WAGNER M., WELLS D.;  
RT "Closely linked H2B genes in the marine copepod, *Tigriopus*  
RT californicus indicate a recent gene duplication or gene conversion  
RT event.";  
RL DNA Seq. 2:387-396(1992).

\*\*\*Note: remainder of annotations omitted.

Query Match 100.0%; Score 69; DB 1; Length 135;  
Best Local Similarity 100.0%; Pred. NO. 2.46e-05;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 40 RYRPGTVAL 48  
QY 1 RYRPGTVAL 9  
|||||||

RESULT 14  
ID H3\_CAEL STANDARD; PRT; 135 AA.  
AC P08898;  
DT 01-NOV-1988 (Rel. 09, Created)  
DT 01-NOV-1990 (Rel. 16, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE HISTONE H3  
GN (HIS-9 OR F22B3.2) AND B0035.10 AND (HIS-6 OR F45F2.13) AND F54E12.1  
GN AND F55G1.2 AND K06C4.5 AND K06C4.13 AND ZK131.2 AND ZK131.3 AND  
GN ZK131.7.  
OS *Caenorhabditis elegans*.  
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;  
OC Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; *Caenorhabditis*.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 89293823.  
RA ROBERTS S.B., EMMONS S.W., CHILDS G.;  
RT "Nucleotide sequences of *Caenorhabditis elegans* core histone genes.  
RT Genes for different histone classes share common flanking sequence  
RT elements.";  
RL J. Mol. Biol. 206:567-577(1989).  
RN [2]  
RP SEQUENCE FROM N.A. (F22B3.2).  
RC STRAIN-BRISTOL N2;  
RA COTTAGE A.;  
RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.  
RN [3]

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
 CC EMBL: M11354; AAA52653.1; -  
 CC EMBL: M11353; AAA52654.1; -  
 CC EMBL: Z48950; CAA88778.1; -  
 CC EMBL: X05855; CAA29288.1; ALT\_SEQ..  
 CC EMBL: X05856; CAA29288.1; JOINED.  
 CC EMBL: X05857; CAA29288.1; JOINED.  
 CC EMBL: X73683; CAA52035.1; -  
 CC EMBL: X13605; CAA31940.1; -  
 CC EMBL: X51897; CAA36179.1; -  
 CC EMBL: M11393; AAA48794.1; -  
 CC EMBL: Y00392; CAA68456.1; -  
 CC EMBL: M17876; AAA29965.1; -  
 CC EMBL: X53822; CAA37819.1; -  
 CC EMBL: X82257; CAA57712.1; -  
 CC EMBL: X81205; CAA57077.1; -  
 CC EMBL: X81206; CAA57078.1; -  
 CC EMBL: X81207; CAA57080.1; -  
 CC EMBL: X81208; CAA57081.1; -  
 CC PIR: A02622; HSHU33.  
 CC PIR: A27501; A27501.  
 CC PIR: S04186; S04186.  
 CC PIR: S10168; S10168.  
 CC PIR: A45941; A45941.  
 CC PIR: JQ1343; JQ1343.  
 CC FLYBASE: FBgn0004828; His3.3B.  
 CC MIM: 601058; -  
 CC MGD: MGI:1101768; H3f3b.  
 CC PROSITE: PS00322; HISTONE\_H3\_1; 1.  
 CC PROSITE: PS00959; HISTONE\_H3\_2; 1.  
 CC PFAM: PF00125; histone; 1.  
 CC Nuclear protein; Chromosomal protein; DNA-binding; Nucleosome core;  
 CC Multigene family.  
 CC INIT\_MET 0  
 CC SEQUENCE 135 AA; 15197 MW; 5B80974A CRC32;

Query Match 100.0%; Score 69; DB 1; Length 135;  
 Best Local Similarity 100.0%; Pred. No. 2.46e-05;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 40 RYRPGTVAL 48  
 QY 1 RYRPGTVAL 9

RESULT 13  
 ID H32\_BOVIN STANDARD; PRT: 135 AA.  
 AC F16105; P02295; P17320; P17269;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE HISTONE H3 (H3.2).  
 OS Bos taurus (Bovine), Mus musculus (Mouse), Gallus gallus (Chicken),  
 OS Cairina moschata (Muscovy duck),  
 OS Ictiobus bubalus (Smallmouth buffalo fish),  
 OS Poroderma africanus (African catfish) (African catfish),  
 OS Xenopus laevis (African clawed frog),  
 OS Oncorhynchus mykiss (Rainbow trout) (Salmo gairdneri),  
 OS Platynereis dumerilii (Dumeril's clam worm),  
 OS Drosophila hydei (fruit fly), Tigrilopus californicus,  
 OS Chironomus thummi thummi (Midge), and  
 OS Urechis caupo (Innkeeper worm) (Spoonworm).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Bovinae; Bos.  
 OC [1]

RP SEQUENCE.  
 RC SPECIES=BOVINE; TISSUE=THYMUS;  
 RX MEDLINE: 72154496.  
 RA MARZLUFF W.F. JR., SANDERS L.A., MILLER D.M., MCCARTY K.S.;  
 RT "Two chemically and metabolically distinct forms of calf thymus  
 RT histone F3.";  
 RL J. Biol. Chem. 247:2026-2033(1972).  
 RN [2]  
 RP SEQUENCE.  
 RC SPECIES=BOVINE;  
 RX MEDLINE: 75095680.  
 RA PARRY L., SMITH E.L.;  
 RT "Histone III. VI. Two forms of calf thymus histone III.";  
 RL J. Biol. Chem. 250:1919-1920(1975).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=MOUSE;  
 RX MEDLINE: 84041477.  
 RA SITTMAN D.B., GRAVES R.A., MARZLUFF W.F.;  
 RT "Structure of a cluster of mouse histone genes.";  
 RL Nucleic Acids Res. 11:6679-6697(1983).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=MOUSE; STRAIN=BALB/C; TISSUE=LIVER;  
 RX MEDLINE: 90067871.  
 RA HURT M.M., CHODCHOY N., MARZLUFF W.F.;  
 RT "The mouse histone H2a.2 gene from chromosome 3.";  
 RL Nucleic Acids Res. 17:8876-8876(1989).  
 RN [5]  
 RP SEQUENCE FROM N.A. (H3.2-221 AND H3.2-614).  
 RC SPECIES=MOUSE;  
 RX MEDLINE: 87112762.  
 RA TAYLOR J.D., WELLMAN S.E., MARZLUFF W.F.;  
 RT "Sequences of four mouse histone H3 genes: implications for evolution  
 RT of mouse histone genes.";  
 RL J. Mol. Evol. 23:242-249(1986).  
 RN [6]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=MOUSE;  
 RX MEDLINE: 91065547.  
 RA GRUBER A., STREIT A., REIST M., BENNINGER P., BOEHNI R.,  
 RT "Structure of a mouse histone-encoding gene cluster.";  
 RL Gene 95:303-304(1990).  
 RN [7]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=CHICKEN;  
 RX MEDLINE: 85215552.  
 RA WANG S.W., ROBINS A.J., D'ANDREA R., WELLS J.R.E.;  
 RT "Inverted duplication of histone genes in chicken and disposition of  
 RT regulatory sequences.";  
 RL Nucleic Acids Res. 13:1369-1387(1985).  
 RN [8]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=CHICKEN;  
 RX MEDLINE: 82195575.  
 RA ENGEL J.D., SUGARMAN B.J., DODGSON J.B.;  
 RT "A chicken histone H3 gene contains intervening sequences.";  
 RL Nature 297:434-436(1982).  
 RN [9]  
 RP SEQUENCE FROM N.A. (H3-II AND H3-III).  
 RC SPECIES=CHICKEN;  
 RX MEDLINE: 91340167.  
 RA NAKAYAMA T.;  
 RT "Nucleotide sequences of two members of the chicken H3  
 RT histone-encoding gene family.";  
 RL Gene 102:289-290(1991).  
 RN [10]  
 RP SEQUENCE FROM N.A. (H3-IV AND H3-V).  
 RC SPECIES=CHICKEN; STRAIN=WHITE LEHORN;  
 RX MEDLINE: 92066488.  
 RA SETOGUCHI Y., NAKAYAMA T.;  
 RT "Nucleotide sequences of new members (H3-IV and H3-V) of the chicken

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; X05222; CAA28851.1; -;  
 DR EMBL; X05223; CAA28852.1; -;  
 DR EMBL; AB004538; BAA21441.1; -;  
 DR EMBL; AL022072; BAA17819.1; -;  
 DR PIR; E27399; HS2P3.  
 DR PROSITE; PS00322; HISTONE\_H3\_1; 1.  
 DR PROSITE; PS00959; HISTONE\_H3\_2; 1.  
 DR PFAM; PF00125; histone; 1.  
 KW Nuclear protein; Chromosomal protein; DNA-binding; Nucleosome core;  
 FT INIT\_MET 0  
 SQ SEQUENCE 135 AA; 15226 MW; 9932E953 CRC32;

Query Match 100.0%; Score 69; DB 1; Length 135;  
 Best Local Similarity 100.0%; Pred. No. 2.46e-05;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 40 RYRPGTVAL 48  
 |||||  
 QY 1 RYRPGTVAL 9

RESULT 12  
 ID H33\_HUMAN STANDARD; PRT; 135 AA.  
 AC P06351; P33155;  
 DT 01-JAN-1988 (Rel. 06, Created)  
 DT 01-JAN-1988 (Rel. 06, Last sequence update)  
 DT 15-DEC-1999 (Rel. 39, Last annotation update)  
 DE HISTONE H3.3 (H3.B) (H3.30).  
 GN H3F3B OR H3H3-3Q OR H3S3.3B.  
 OS Homo sapiens (Human); Mus musculus (Mouse); Rattus norvegicus (Rat);  
 OS Oryctolagus cuniculus (Rabbit); Gallus gallus (Chicken);  
 OS Spisula solidissima (Atlantic surf-clam);  
 OS Drosophila melanogaster (Fruit fly); and Drosophila hydei (Fruit fly).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 RN [1]  
 RC SEQUENCE FROM N.A.  
 RP SPECIES=HUMAN;  
 RX MEDLINE; 85190590.  
 RA WELLS D., KEDES L.;  
 RT "Structure of a human histone cDNA: evidence that basally expressed  
 RT histone genes have intervening sequences and encode polyadenylated  
 RT mRNAs";  
 RL Proc. Natl. Acad. Sci. U.S.A. 82:2834-2838(1985).  
 RN [2]  
 RC SEQUENCE FROM N.A.  
 RP SPECIES=HUMAN;  
 RX MEDLINE; 87174815.  
 RA WELLS D., HOFFMAN D., KEDES L.;  
 RT "Unusual structure, evolutionary conservation of non-coding sequences  
 RT and numerous pseudogenes characterize the human H3.3 histone  
 RT multigene family";  
 RL Nucleic Acids Res. 15:2871-2889(1987).  
 RN [3]  
 RC SEQUENCE FROM N.A.  
 RP SPECIES=HUMAN;  
 RX MEDLINE; 96163879.  
 RA ALBIG W., BRAMLAGE B., GRUBER K., KLOBECK H.-G., KUNZ J., DOENECKE D.;  
 RT "The human replacement histone H3.3B gene (H3F3B)";  
 RL Genomics 30:264-272(1995).  
 RN [4]  
 RC PARTIAL SEQUENCE.  
 RP SPECIES=HUMAN;  
 RX MEDLINE; 82075746.

RA ONE Y., IWAI K.;  
 RT "Human spleen histone H3. Isolation and amino acid sequence.";  
 RL J. Biochem. 90:1205-1211(1981).  
 RN [5]  
 RC SEQUENCE FROM N.A.  
 RP SPECIES=MOUSE;  
 RX MEDLINE; 89240011.  
 RA HRABA-RENEVEY S., KRESS M.;  
 RT "Expression of a mouse replacement histone H3.3 gene with a highly  
 RT conserved 3' noncoding region during SV40- and polyoma-induced Go to  
 RT S-phase transition.";  
 RL Nucleic Acids Res. 17:2449-2461(1989).  
 RN [6]  
 RC SEQUENCE FROM N.A.  
 RP SPECIES=RAT; STRAIN=SPRAGUE-DAWLEY; TISSUE=BRAIN;  
 RA DI LIEGRO I.;  
 RL Submitted (JUN-1993) to the EMBL/GenBank/DBJ databases.  
 RN [7]  
 RC SEQUENCE FROM N.A.  
 RP SPECIES=RABBIT;  
 RX MEDLINE; 90272438.  
 RA CHALMERS M., WELLS D.;  
 RT "Extreme sequence conservation characterizes the rabbit H3.3A histone  
 RT cDNA";  
 RL Nucleic Acids Res. 18:3075-3075(1990).  
 RN [8]  
 RC SEQUENCE FROM N.A.  
 RP SPECIES=CHICKEN;  
 RX MEDLINE; 85295962.  
 RA BRUSH D., DODGSON J.B., CHOI O.R., WILKINS STEVENS P., ENGEL J.D.;  
 RT "Replacement variant histone genes contain intervening sequences";  
 RL Mol. Cell. Biol. 5:1307-1317(1985).  
 RN [9]  
 RC SEQUENCE FROM N.A.  
 RP SPECIES=CHICKEN; STRAIN=WHITE LEGHORN; TISSUE=LIVER;  
 RX MEDLINE; 87316866.  
 RA DODGSON J.B., YAMAMOTO M., ENGEL J.D.;  
 RT "Chicken histone H3.3B cDNA sequence confirms unusual 3' UTR  
 RT structure";  
 RL Nucleic Acids Res. 15:6294-6294(1987).  
 RN [10]  
 RC SEQUENCE FROM N.A.  
 RP SPECIES=S.SOLIDISSIMA;  
 RX MEDLINE; 87305176.  
 RA SWENSON K.I., BORGESE N., PIETRINI G., RUDERMAN J.V.;  
 RT "Three translationally regulated mRNAs are stored in the cytoplasm of  
 RT clam oocytes";  
 RL Dev. Biol. 123:10-16(1987).  
 RN [11]  
 RC SEQUENCE FROM N.A.  
 RP SPECIES=D.MELANOGASTER;  
 RX MEDLINE; 92084129.  
 RA FRETZIN S., ALLAN B.D., VAN DAAL A., ELGIN S.C.R.;  
 RT "A Drosophila melanogaster H3.3 cDNA encodes a histone variant  
 RT identical with the vertebrate H3.3";  
 RL Gene 107:341-342(1991).  
 RN [12]  
 RC SEQUENCE FROM N.A.  
 RP SPECIES=D.MELANOGASTER, AND D.HYDEI; STRAIN=TUBINGEN;  
 RX MEDLINE; 96023949.  
 RA AKHMANOVA A.S., BINDELS P.S.T., XU J., MIEDEMA K., KREMER H.,  
 RA HENNIG W.;  
 RT "Structure and expression of histone H3.3 genes in Drosophila  
 RT melanogaster and Drosophila hydei";  
 RL Genome 38:586-600(1995).  
 CC -1- FUNCTION: HISTONE H3, ALONG WITH HISTONE H4, PLAYS A CENTRAL ROLE  
 CC IN NUCLEOSOME FORMATION.  
 CC -1- SUBUNIT: THE NUCLEOSOME IS AN OCTAMER CONTAINING TWO MOLECULES OF  
 CC H2A, H2B, H3, AND H4; WHICH WRAP APPROXIMATIVELY 146 BP OF DNA.  
 CC -1- MISCELLANEOUS: THIS HISTONE IS THE PREDOMINANT FORM IN NONDIVIDING  
 CC CELLS.  
 CC -1- SIMILARITY: BELONGS TO THE HISTONE H3 FAMILY.  
 CC -----

```
DT 01-APR-1990 (Rel. 14, Last sequence update)
DE 15-JUL-1999 (Rel. 38, Last annotation update)
GN HISTONE H3.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-AK-194.
RX MEDLINE; 89098383.
RA MATSUO Y., YAMAZAKI T.;
RT "tRNA derived insertion element in histone gene repeating unit of
RT Drosophila melanogaster."
RL Nucleic Acids Res. 17:225-238(1989).
RN [2]
RP SEQUENCE FROM N.A.
RA GOLDBERG M.L.;
RL Thesis (1979), University of Stanford, U.S.A.
CC -1- FUNCTION: HISTONE H3, ALONG WITH HISTONE H4, PLAYS A CENTRAL ROLE
CC IN NUCLEOSOME FORMATION.
CC -1- SUBUNIT: THE NUCLEOSOME IS AN OCTAMER CONTAINING TWO MOLECULES OF
CC H2A, H2B, H3, AND H4; WHICH WRAP APPROXIMATELY 146 BP OF DNA.
CC -1- DEVELOPMENTAL STAGE: THIS HISTONE IS EXPRESSED DURING LATE
CC EMBRYONIC DEVELOPMENT.
CC -1- SIMILARITY: BELONGS TO THE HISTONE H3 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X03952; CAA27582.1;
DR PIR; A02629; HSUR3P.
DR PROSITE; PS00322; HISTONE_H3_1; 1.
DR PROSITE; PS00959; HISTONE_H3_2; 1.
DR PFAM; PF00125; histone; 1.
KW Nuclear protein; Chromosomal protein; DNA-binding; Nucleosome core;
KW Embryo.
FT INIT_MET 0
SQ SEQUENCE 135 AA; 15370 MW; F9EBFB65 CRC32;
-----
Query Match 100.0%; Score 69; DB 1; Length 135;
Best Local Similarity 100.0%; Pred. No. 2.46e-05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
-----
Db 40 RYRPGTVAL 48
QY 1 RYRPGTVAL 9
-----
RESULT 11
ID H31_SCHPO STANDARD; PRT; 135 AA.
AC P09988;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE HISTONE H3.1/H3.2.
GN (HHT1 OR SPBC8D2.04 OR PI060) AND HHT2.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Archiascomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
RN [1]
RP SEQUENCE FROM N.A. (HHT1).
RX MEDLINE; 86135992.
RA MATSUMOTO S., YANAGIDA M.;
RT "Histone gene organization of fission yeast: a common upstream
RT sequence."
RL EMBO J. 4:3531-3538(1985).
RN [2]
RP SEQUENCE FROM N.A. (HHT1).
RC STRAIN-972;
RA KUSHIDA N., YAMAZAKI S., TANAKA T., JINNO K., HAIKAWA Y., YAMAZAKI J.,
RA YAMAMOTO S., SEKINE M., OGUCHI A., NAGAI Y., SAKAI M., AOKI K.,
RA OGURA K., OTSUKA R., KUDOH Y., YANAGIDA M., MACHIDA M., ZHANG M.Q.;
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A. (HHT1).
RC STRAIN-972;
RA LYNE M., RAJANDREAM M.A., BARRELL B.G., LAUBER J., HILBERT H.,
RA DUESTERHOEF A.;
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: HISTONE H3, ALONG WITH HISTONE H4, PLAYS A CENTRAL ROLE
CC IN NUCLEOSOME FORMATION.
CC -1- SUBUNIT: THE NUCLEOSOME IS AN OCTAMER CONTAINING TWO MOLECULES OF
CC H2A, H2B, H3, AND H4; WHICH WRAP APPROXIMATELY 146 BP OF DNA.
CC -1- SIMILARITY: BELONGS TO THE HISTONE H3 FAMILY.
CC -----
```

```
QY 1 RYRPGTVAL 9
RESULT 8
ID H31_HUMAN STANDARD; PRT; 135 AA.
AC P16106; P02295; P02296;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE HISTONE H3.1
OS Homo sapiens (Human), Bos taurus (Bovine), and Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=HUMAN;
RX MEDLINE; 86242753.
RA MARASHI F., HELMS S., SHIELDS A., SILVERSTEIN S., GREENSPAN D.S.,
RA STEIN G., STEIN J.;
RT "Enhancer-facilitated expression of prokaryotic and eukaryotic genes
RT using human histone gene 5' regulatory sequences.";
RL Biochem. Cell Biol. 64:277-289(1986).
RN [3]
RP SEQUENCE FROM N.A.
RC SPECIES=HUMAN;
RX MEDLINE; 92009931.
RA ALBIG W., KARDALINO E., DRABENT B., ZIMMER A., DOENECKE D.;
RT "Isolation and characterization of two human H1 histone genes within
RT clusters of core histone genes.";
RL Genomics 10:940-948(1991).
RN [4]
RP PARTIAL SEQUENCE.
RC SPECIES=HUMAN;
RX MEDLINE; 82075746.
RA OHE Y., IWAI K.;
RT "Human spleen histone H3. Isolation and amino acid sequence.";
RL J. Biochem. 90:1205-1211(1981).
RN [5]
RP SEQUENCE.
RC SPECIES=BOVINE;
RX MEDLINE; 73166574.
RA DELANGE R.J., HOOPER J.A., SMITH E.L.;
RT "Histone 3. 3. Sequence studies on the cyanogen bromide peptides;
RT complete amino acid sequence of calf thymus histone 3.";
RL J. Biol. Chem. 248:3261-3274(1973).
RN [6]
RP PARTIAL SEQUENCE.
RC SPECIES=BOVINE;
RX MEDLINE; 73166572.
RA DELANGE R.J., SMITH E.L.;
RT "Histone 3. I. Isolation and sequences of the tryptic peptides from
RT the maleylated calf thymus protein.";
RL J. Biol. Chem. 248:3248-3254(1973).
RN [7]
RP PARTIAL SEQUENCE.
RC SPECIES=BOVINE;
RX MEDLINE; 73166573.
RA HOOPER J.A., SMITH E.L.;
RT "Histone 3. II. Isolation and sequences of chymotryptic peptides from
RT calf thymus histone 3.";
RL J. Biol. Chem. 248:3255-3260(1973).
RN [8]
RP SEQUENCE FROM N.A.
RC SPECIES=MOUSE; STRAIN=CD-1; TISSUE=TESTIS;
RX MEDLINE; 90067856.
RA KOSCISSA U., DOENECKE D.;
RT "Nucleotide sequences of mouse histone genes H2A and H3.1.";
RL Nucleic Acids Res. 17:8861-8861(1989).
RN [9]
RP SEQUENCE FROM N.A.
RC SPECIES=MOUSE;
RX MEDLINE; 84041477.
RA SITTMAN D.B., GRAVES R.A., MARZLUFF W.F.;
RT "Structure of a cluster of mouse histone genes.";
RL Nucleic Acids Res. 11:6679-6697(1983).
RN [10]
RP SEQUENCE FROM N.A. (H3.1-221 AND H3.1-291).
RC SPECIES=MOUSE;
RX MEDLINE; 87112762.
RA TAYLOR J.D., WELLMAN S.E., MARZLUFF W.F.;
RT "Sequences of four mouse histone H3 genes: Implications for evolution
RT of mouse histone genes.";
RL J. Mol. Evol. 23:242-249(1986).
CC -I- FUNCTION: HISTONE H3, ALONG WITH HISTONE H4, PLAYS A CENTRAL ROLE
CC IN NUCLEOSOME FORMATION.
CC -I- SUBUNIT: THE NUCLEOSOME IS AN OCTAMER CONTAINING TWO MOLECULES OF
CC H2A, H2B, H3, AND H4; WHICH WRAP APPROXIMATIVELY 146 BP OF DNA.
CC -I- SIMILARITY: BELONGS TO THE HISTONE H3 FAMILY.
CC
CC THIS SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X00090; CAA24952.1; -
CC EMBL; M26150; AAA52651.1; -
CC EMBL; M60746; AAA63185.1; -
CC EMBL; X57128; CAA40407.1; -
CC EMBL; X16496; CAA34512.1; -
CC EMBL; X01684; CAA35839.1; -
CC EMBL; M32460; AAA37811.1; -
CC EMBL; M32462; AAA37813.1; -
CC PIR; A02623; HSHU3.
CC PIR; A02624; HSB03.
CC PIR; A40335; A40335.
CC PIR; S06755; S06755.
CC PIR; S28528; S28528.
CC MIM; 142780; -
CC PROSITE; PS00322; HISTONE_H3_1; 1.
CC PROSITE; PS00959; HISTONE_H3_2; 1.
CC PFAM; PF00125; histone; 1.
CC
CC Nuclear protein; Chromosomal protein; DNA-binding; Nucleosome core;
CC Multigene family; Acetylation; Methylation.
FT INIT_MET 0 0
FT MOD_RES 9 9 METHYLATION.
FT MOD_RES 14 14 ACETYLATION.
FT MOD_RES 23 23 ACETYLATION.
FT MOD_RES 27 27 METHYLATION.
FT MOD_RES 36 36 METHYLATION.
FT CONFLICT 134 134 MISSING (IN REF. 2).
SQ SEQUENCE 135 AA; 15273 MW; EC3247C6 CRC32;
Query Match 100.0%; Score 69; DB 1; Length 135;
Best Local Similarity 100.0%; Pred. No. 2.46e-05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 40 RYRPGTVAL 48
| | | | |
QY 1 RYRPGTVAL 9
RESULT 9
ID H3_DROME STANDARD; PRT; 135 AA.
AC P02299;
DT 21-JUL-1986 (Rel. 01, Created)
```



RESULT 6  
ID H32\_MSDA STANDARD; PRT; 135 AA.  
AC P11105;  
DT 01-JUL-1989 (Rel. 11, Created)  
DT 01-NOV-1990 (Rel. 16, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE HISTONE H3.2, MINOR.  
OS Medicago sativa (alfalfa), Arabidopsis thaliana (Mouse-ear cress), and  
OS Lolium temulentum.  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC eukaryophytes; Spermatophyta; Magnoliophyta; eudicotyledons;  
OC core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae;  
OC Medicago.  
RN [1]  
RP SPECIES FROM N.A.  
RC SPECIES-M.SATIYA; STRAIN-CV. CHIEF, AND CV. REGEN S;  
RA ROBERTSON A.J.;  
RL Thesis (1994), University of Missouri / Kansas City, U.S.A.  
RN [2]  
RP SPECIES OF 17-135 FROM N.A.  
RC SPECIES-M.SATIYA; STRAIN-CV. REGEN S;  
RX MEDLINE; 89263717.  
RA WU S.C., GYERREY J., DUDITS D.;  
RT "Polyadenylated H3 histone transcripts and H3 histone variants in  
RT alfalfa."  
RL Nucleic Acids Res. 17:3057-3063(1989).  
RN [3]  
RP SPECIES OF 1-44 AND 84-115, ACETYLATION, AND METHYLATION.  
RC SPECIES-M.SATIYA; STRAIN-CV. R4;  
RX MEDLINE; 91009145.  
RA WATERBORG J.H.;  
RT "Sequence analysis of acetylation and methylation in two histone H3  
RT variants of alfalfa."  
RL J. Biol. Chem. 265:17157-17161(1990).  
RN [4]  
RP SPECIES FROM N.A.  
RC SPECIES-A.THALIANA; STRAIN-CV. COLUMBIA;  
RX MEDLINE; 92277663.  
RA CHAUBET N., CLEMENT B., GIGOT C.;  
RT "Genes encoding a histone H3.3-like variant in Arabidopsis contain  
RT intervening sequences."  
RL J. Mol. Biol. 225:569-574(1992).  
RN [5]  
RP SPECIES FROM N.A.  
RC SPECIES-L.TEMULENTUM;  
RA FREEMAN D.R., OUGHAM H.J.;  
RL Submitted (JUL-1994) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: HISTONE H3, ALONG WITH HISTONE H4, PLAYS A CENTRAL ROLE  
CC IN NUCLEOSOME FORMATION.  
CC -1- SUBUNIT: THE NUCLEOSOME IS AN OCTAMER CONTAINING TWO MOLECULES OF  
CC H2A, H2B, H3, AND H4; WHICH WRAP APPROXIMATELY 146 BP OF DNA.  
CC -1- SIMILARITY: BELONGS TO THE HISTONE H3 FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; U09458; AAB49538.1; -  
CC EMBL; U09460; AAB36493.1; -  
CC EMBL; U09461; AAB36494.1; -  
CC EMBL; U09464; AAB36497.1; -  
CC EMBL; U09465; AAB36498.1; -  
CC EMBL; X13676; CAA31967.1; -  
CC EMBL; X60429; CAA42958.1; -  
CC EMBL; X60429; CAA42957.1; -  
CC EMBL; X79714; CAA56153.1; -  
CC PIR; B38309; B38309.  
CC PIR; S04521; S04521.

PIR; S24346; S24346.  
DR PROSITE; PS00322; HISTONE\_H3\_1; 1.  
DR PROSITE; PS00959; HISTONE\_H3\_2; 1.  
DR PFAM; PF00125; histone; 1.  
KW Nuclear protein; Chromosomal protein; DNA-binding; Nucleosome core;  
KW Multigene family; Acetylation; Methylation.  
FT INIT\_MET 0 0  
FT MOD\_RES 4 4 METHYLATION.  
FT MOD\_RES 9 9 ACETYLATION.  
FT MOD\_RES 14 14 METHYLATION.  
FT MOD\_RES 14 14 ACETYLATION.  
FT MOD\_RES 18 18 METHYLATION.  
FT MOD\_RES 18 18 ACETYLATION.  
FT MOD\_RES 23 23 METHYLATION.  
FT MOD\_RES 23 23 ACETYLATION.  
FT MOD\_RES 27 27 METHYLATION.  
SQ SEQUENCE 135 AA; 15275 MW; F32F962A CRC32;  
Query Match 100.0%; Score 69; DB 1; Length 135;  
Best Local Similarity 100.0%; Pred. No. 2.46e-05;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Db 40 RYRPGTVAL 48  
QY 1 RYRPGTVAL 9  
RESULT 7  
ID H33\_CAEEL STANDARD; PRT; 135 AA.  
AC Q10453;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE HISTONE H3.3.  
GN F45E1.6.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;  
OC Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RA WATERSTON R.;  
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: HISTONE H3, ALONG WITH HISTONE H4, PLAYS A CENTRAL ROLE  
CC IN NUCLEOSOME FORMATION.  
CC -1- SUBUNIT: THE NUCLEOSOME IS AN OCTAMER CONTAINING TWO MOLECULES OF  
CC H2A, H2B, H3, AND H4; WHICH WRAP APPROXIMATELY 146 BP OF DNA.  
CC -1- SIMILARITY: BELONGS TO THE HISTONE H3 FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; U28732; AAB04902.1; -  
CC EMBL; F45E1.6; CE10488.  
DR PROSITE; PS00322; HISTONE\_H3\_1; 1.  
DR PROSITE; PS00959; HISTONE\_H3\_2; 1.  
DR PFAM; PF00125; histone; 1.  
KW Nuclear protein; Chromosomal protein; DNA-binding; Nucleosome core;  
KW Multigene family.  
FT INIT\_MET 0 0 BY SIMILARITY.  
FT MOD\_RES 4 4 METHYLATION.  
SQ SEQUENCE 135 AA; 15211 MW; 2F167D92 CRC32;  
Query Match 100.0%; Score 69; DB 1; Length 135;  
Best Local Similarity 100.0%; Pred. No. 2.46e-05;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Db 40 RYRPGTVAL 48  
QY 1 RYRPGTVAL 9

```

RN [4]
RP SEQUENCE FROM N.A.
RC SPECIES=P.LIVIDUS;
RA SPINELLI G.;
RL Submitted (FEB-1989) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RC SPECIES=P.LIVIDUS, AND S.DROEBACHIENSIS;
RA BUSSELMING M., RUSCONI S., BIRNSTIEL M.L.;
RT "An unusual evolutionary behaviour of a sea urchin histone gene cluster.";
RL EMBO J. 1:27-33(1982).
RN [6]
RP SEQUENCE FROM N.A.
RC SPECIES=L.PICTUS;
RX MEDLINE: 84216304;
RA ROBERTS S.B., WEISSER K.E., CHILDS G.;
RT "Sequence comparisons of non-allelic late histone genes and their early stage counterparts. Evidence for gene conversion within the sea urchin late stage gene family.";
RL J. Mol. Biol. 174:647-662(1984).
RN [7]
RP SEQUENCE FROM N.A.
RC SPECIES=D.IMBRICATA, P.BREVISPINUS, AND P.OCHRACEUS; TISSUE=SPERM;
RX MEDLINE: 88259237;
RA BANFIELD D.C.D., HONDA B.M., SMITH M.J.;
RT "Histone genes in three sea star species: cluster arrangement, transcriptional polarity, and analyses of the flanking regions of H3 and H4 genes.";
RL J. Mol. Evol. 27:36-44(1988).
RN [8]
RP SEQUENCE FROM N.A.
RC SPECIES=P.OCHRACEUS, P.BREVISPINUS, P.HELICANTHOIDES, AND S.STIMPSONI;
RA WU Y., KOWBEL D., SMITH M.J.;
RL Submitted (JUL-1990) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: HISTONE H3, ALONG WITH HISTONE H4, PLAYS A CENTRAL ROLE IN NUCLEOSOME FORMATION.
CC -!- SUBUNIT: THE NUCLEOSOME IS AN OCTAMER CONTAINING TWO MOLECULES OF H2A, H2B, H3, AND H4; WHICH WRAP APPROXIMATELY 146 BP OF DNA.
CC -!- DEVELOPMENTAL STAGE: THIS HISTONE IS EXPRESSED DURING LATE EMBRYONIC DEVELOPMENT.
CC -!- SIMILARITY: BELONGS TO THE HISTONE H3 FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC
DR EMBL; J01181; AAB59206.1;
DR EMBL; X01345; CAA25632.1;
DR EMBL; V01143; CAA24375.1;
DR EMBL; V01144; CAA24382.1; ALT_SEQ.
DR EMBL; M10558; AAA30026.1;
DR EMBL; M25281; AAA65843.1;
DR EMBL; X07505; CAA30388.1;
DR EMBL; M36919; AAA75395.1;
DR EMBL; M36920; AAA29441.1;
DR EMBL; M36921; AAA30053.1;
DR EMBL; X00628; CAA25262.1;
DR EMBL; X00593; CAA25242.1;
DR EMBL; X07504; CAA30387.1;
DR EMBL; X07503; CAA30386.1;
DR EMBL; X54112; CAA38050.1;
DR EMBL; X54113; CAA38052.1;
DR EMBL; X54114; CAA38054.1;
DR EMBL; X54115; CAA38056.1;
DR PIR; A02628; HSUR3M.
DR PIR; S01196; S01196.
DR PIR; S01197; S01197.
DR PIR; S01198; S01198.

```

```

DR PIR; S20667; S20667.
DR PIR; S20671; S20671.
DR PIR; S20678; S20678.
DR PIR; S20669; S20669.
DR PROSITE; PS00322; HISTONE_H3_1; 1.
DR PROSITE; PS00959; HISTONE_H3_2; 1.
DR PFAM; PF00125; histone; 1.
KW Nuclear protein; Chromosomal protein; DNA-binding; Nucleosome core; Embryo.
FT INIT_MET 0
SQ SEQUENCE 135 AA; 15271 MW; 6AD6F728 CRC32;
Query Match 100.0%; Score 69; DB 1; Length 135;
Best Local Similarity 100.0%; Pred. No. 2.46e-05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 40 RYRPGTVAL 48
QY 1 RYRPGTVAL 9
|||||
RESULT 5
ID H32_XENLA STANDARD; PRT; 135 AA.
AC P03302;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DE HISTONE H3.2
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Amphibia;
OC Eukaryota; Anura; Mesobatrachia; Pipidae; Xenopodinae;
OC Xenopus.
RN [1]
RP SEQUENCE FROM N.A. (GENE CLUSTER X1H1).
RX MEDLINE: 82095633;
RA MOORMAN A.F.M., DE BOER P.A.J., DE LAAF R.T.M., VAN DONGEN W.M.A.M., DESTREE O.H.J.;
RT "Primary structure of the histone H3 and H4 genes and their flanking sequences in a minor histone gene cluster of Xenopus laevis.";
RL FEBS Lett. 136:45-52(1981).
CC -!- FUNCTION: HISTONE H3, ALONG WITH HISTONE H4, PLAYS A CENTRAL ROLE IN NUCLEOSOME FORMATION.
CC -!- SUBUNIT: THE NUCLEOSOME IS AN OCTAMER CONTAINING TWO MOLECULES OF H2A, H2B, H3, AND H4; WHICH WRAP APPROXIMATELY 146 BP OF DNA.
CC -!- SIMILARITY: BELONGS TO THE HISTONE H3 FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC
DR EMBL; J00982; -; NOT_ANNOTATED_CDS.
DR EMBL; J00984; -; NOT_ANNOTATED_CDS.
DR PIR; A02634; HXSL32.
DR PROSITE; PS00322; HISTONE_H3_1; 1.
DR PROSITE; PS00959; HISTONE_H3_2; 1.
DR PFAM; PF00125; histone; 1.
KW Nuclear protein; Chromosomal protein; DNA-binding; Nucleosome core; Multigene family.
FT INIT_MET 0
SQ SEQUENCE 135 AA; 15356 MW; 9DA3E094 CRC32;
Query Match 100.0%; Score 69; DB 1; Length 135;
Best Local Similarity 100.0%; Pred. No. 2.46e-05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 40 RYRPGTVAL 48
QY 1 RYRPGTVAL 9
|||||

```

```

RESULT 2
ID H3_CHLRE STANDARD; PRT; 134 AA.
AC P05564;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE HISTONE H3.
GN H3.
OS Chlamydomonas reinhardtii.
OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
OC Chlamydomonadaceae; Chlamydomonas.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=137;
RX MEDLINE: 96017782.
RA WALTHER Z., HALL J.L.;
RT structure.";
RL Nucleic Acids Res. 23:3756-3763(1995).
CC -!- FUNCTION: HISTONE H3, ALONG WITH HISTONE H4, PLAYS A CENTRAL ROLE
CC IN NUCLEOSOME FORMATION.
CC -!- SUBUNIT: THE NUCLEOSOME IS AN OCTAMER CONTAINING TWO MOLECULES OF
CC H2A, H2B, H3, AND H4; WHICH WRAP APPROXIMATIVELY 146 BP OF DNA.
CC -!- SIMILARITY: BELONGS TO THE HISTONE H3 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L41841; AAA99965.1; .
DR PROSITE; PS00322; HISTONE_H3_1; 1.
DR PROSITE; PS00959; HISTONE_H3_2; 1.
DR PFAM; PF00125; histone; 1.
DR Nucleic Acid Protein; Chromosomal protein; DNA-binding; Nucleosome core.
FT INIT_MET 0
SQ SEQUENCE 134 AA; 15129 MW; 281606B6 CRC32;

Query Match 100.0%; Score 69; DB 1; Length 134;
Best Local Similarity 100.0%; Pred. No. 2.46e-05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 39 RYRPGTVAL 47
QY 1 RYRPGTVAL 9

RESULT 3
ID H3_VOLCA STANDARD; PRT; 134 AA.
AC P08437;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE HISTONE H3.
GN H3-I AND H3-II.
OS Volvox carterii.
OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
OC Volvocaceae; Volvox.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HK10.
RX MEDLINE: 88234003.
RA MUELLER K., SCHMITT R.;
RT "Histone genes of Volvox carteri: DNA sequence and organization of
RT two H3-H4 gene loci.";
RL Nucleic Acids Res. 16:4121-4136(1988).
CC -!- FUNCTION: HISTONE H3, ALONG WITH HISTONE H4, PLAYS A CENTRAL ROLE
CC IN NUCLEOSOME FORMATION.
CC -----
DR EMBL; L41841; AAA99965.1; .
DR PROSITE; PS00322; HISTONE_H3_1; 1.
DR PROSITE; PS00959; HISTONE_H3_2; 1.
DR PFAM; PF00125; histone; 1.
DR Nucleic Acid Protein; Chromosomal protein; DNA-binding; Nucleosome core.
FT INIT_MET 0
SQ SEQUENCE 134 AA; 15129 MW; 281606B6 CRC32;

Query Match 100.0%; Score 69; DB 1; Length 134;
Best Local Similarity 100.0%; Pred. No. 2.46e-05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 39 RYRPGTVAL 47
QY 1 RYRPGTVAL 9

```

```

CC -!- SUBUNIT: THE NUCLEOSOME IS AN OCTAMER CONTAINING TWO MOLECULES OF
CC H2A, H2B, H3, AND H4; WHICH WRAP APPROXIMATIVELY 146 BP OF DNA.
CC -!- SIMILARITY: BELONGS TO THE HISTONE H3 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X05963; CAA30035.1; .
DR EMBL; X05964; CAA30037.1; .
DR PIR; S00940; S00940.
DR PROSITE; PS00322; HISTONE_H3_1; 1.
DR PROSITE; PS00959; HISTONE_H3_2; 1.
DR PFAM; PF00125; histone; 1.
DR Nucleic Acid Protein; Chromosomal protein; DNA-binding; Nucleosome core.
FT INIT_MET 0
SQ SEQUENCE 134 AA; 15179 MW; 0FE12547 CRC32;

Query Match 100.0%; Score 69; DB 1; Length 134;
Best Local Similarity 100.0%; Pred. No. 2.46e-05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 39 RYRPGTVAL 47
QY 1 RYRPGTVAL 9

RESULT 4
ID H3_PSAMI STANDARD; PRT; 135 AA.
AC P02298; P05320; P05321; P05322;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE HISTONE H3, EMBRYONIC.
OS Psammecchinus miliaris (Sand sea urchin).
OS Paracentrotus lividus (Common sea urchin).
OS Strongylocentrotus droebachiensis (Sea urchin).
OS Lytechinus pictus (Painted sea urchin).
OS Dermasterias imbricata (Sea star), Pisaster brevispinus (Sea star),
OS Pisaster ochraceus (Sea star),
OS Pycnopodia helianthoides (Sea star), and
OS Solaster stimpsoni (Sea star).
OC Eukaryota; Metazoa; Echinodermata; Echinozoa; Echinoidea;
OC Euechinozoa; Echinacea; Echinoida; Echinidae; Psammecchinus.
RN [1]
RP SEQUENCE FROM N.A. (CLONE H22).
RC SPECIES=P.MILIARIS;
RX MEDLINE: 79001915.
RA SCHAFFNER W., KUNZ G., DAETWYLER H., TELFORD J., SMITH H.O.,
RA BIRNSTIEL M.L.;
RT "Genes and spacers of cloned sea urchin histone DNA analyzed by
RT sequencing.";
RL Cell 14:655-671(1978).
RN [2]
RP SEQUENCE FROM N.A. (CLONE H22).
RC SPECIES=P.MILIARIS;
RA BIRNSTIEL M.L., PORTMANN R., BUSSLINGER M., SCHAFFNER W., PROBST E.,
RA KRESSMANN A.;
RT "Functional organization of the histone genes in the sea urchin
RT Psammecchinus: a progress report.";
RL Alfred Benzon Symp. 13:117-132(1979).
RN [3]
RP SEQUENCE FROM N.A. (CLONE H19).
RC SPECIES=P.MILIARIS;
RX MEDLINE: 81076674.
RA BUSSLINGER M., PORTMANN R., IRMINGER J.C., BIRNSTIEL M.L.;
RT "Ubiquitous and gene-specific regulatory 5' sequences in a sea urchin
RT histone DNA clone coding for histone protein variants.";
RL Nucleic Acids Res. 8:957-977(1980).

```

\*\*\*\*\*  
W P S R E H  
\*\*\*\*\*  
(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Sat Apr 15 00:01:48 2000; MasPar time 4.80 Seconds  
Tabular output not generated. 56.050 Million cell updates/sec

Title: >US-08-452-843-11  
Description: (1-9) from US08452843.pep  
Perfect Score: 69  
Sequence: 1 RYRPGTVAL 9

Scoring table: PAM 150  
Gap 15

Searched: 82229 seqs, 29864866 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: swiss-prot38  
1:swissprot

Statistics: Mean 24.183; Variance 25.056; scale 0.965

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	69	100.0	113	1	H38_STYLE HISTONE H3-8 (FRAGMENT	2.46e-05
2	69	100.0	134	1	H3_CHLRE HISTONE H3.	2.46e-05
3	69	100.0	134	1	H3_VOLCA HISTONE H3.	2.46e-05
4	69	100.0	135	1	H3_PSAMI HISTONE H3, EMBRYONIC.	2.46e-05
5	69	100.0	135	1	H32_XENLA HISTONE H3.2.	2.46e-05
6	69	100.0	135	1	H32_MEESA HISTONE H3.2, MINOR.	2.46e-05
7	69	100.0	135	1	H33_CAEEL HISTONE H3.3.	2.46e-05
8	69	100.0	135	1	H31_HUMAN HISTONE H3.1.	2.46e-05
9	69	100.0	135	1	H31_DROME HISTONE H3.	2.46e-05
10	69	100.0	135	1	H3_STRPU HISTONE H3, EMBRYONIC.	2.46e-05
11	69	100.0	135	1	H31_SCHPO HISTONE H3.1/H3.2.	2.46e-05
12	69	100.0	135	1	H33_HUMAN HISTONE H3.3 (H3.B) (H	2.46e-05
13	69	100.0	135	1	H32_BOVIN HISTONE H3 (H3.2).	2.46e-05
14	69	100.0	135	1	H3_CAEEL HISTONE H3.	2.46e-05
15	69	100.0	135	1	H3_ACRFO HISTONE H3.	2.46e-05
16	64	92.8	112	1	H34_STYLE HISTONE H3-4 (FRAGMENT	5.92e-04
17	64	92.8	114	1	H31_STYLE HISTONE H3-1 (FRAGMENT	5.92e-04
18	64	92.8	114	1	H33_STYLE HISTONE H3-3 (FRAGMENT	5.92e-04
19	64	92.8	114	1	H35_STYLE HISTONE H3-5 (FRAGMENT	5.92e-04
20	64	92.8	114	1	H39_STYLE HISTONE H3-2 (FRAGMENT	5.92e-04
21	64	92.8	114	1	H36_STYLE HISTONE H3-6 (FRAGMENT	5.92e-04
22	64	92.8	114	1	H37_STYLE HISTONE H3-7 (FRAGMENT	5.92e-04
23	64	92.8	135	1	H31_TETPY HISTONE H3.1.	5.92e-04

24	64	92.8	135	1	H3_MAIZE HISTONE H3.	5.92e-04
25	64	92.8	135	1	H3_EMENI HISTONE H3.	5.92e-04
26	64	92.8	135	1	H32_ORYSA HISTONE H3.	5.92e-04
27	64	92.8	135	1	H3_ENCAL HISTONE H3.	5.92e-04
28	64	92.8	135	1	H3_PEA HISTONE H3.	5.92e-04
29	64	92.8	135	1	H3_YEAST HISTONE H3.	5.92e-04
30	64	92.8	135	1	H3_NEUCR HISTONE H3.	5.92e-04
31	62	89.9	134	1	H34_MOUSE HISTONE H3.4 (EMBRYONI	2.04e-03
32	60	87.0	135	1	H34_CAIMO HISTONE H3.4.	6.89e-03
33	60	87.0	135	1	H33_SCHPO HISTONE H3.3.	6.89e-03
34	59	85.5	135	1	H33_TETTH HISTONE H3.3 (HV2).	1.26e-02
35	59	85.5	135	1	H32_TETPY HISTONE H3.2.	1.26e-02
36	56	81.2	413	1	NCAP_IHNV NUCLEOCAPSID PROTEIN (	7.31e-02
37	56	81.2	530	1	KPY1_FELCA PYRUVATE KINASE, M1 IS	7.31e-02
38	54	78.3	135	1	ULA6_HCMVA HYPOTHETICAL PROTEIN U	2.29e-01
39	53	76.8	274	1	PSBO_SYNY3 PHOTOSYSTEM II MANGANE	4.01e-01
40	51	73.9	235	1	YA99_METJA HYPOTHETICAL PROTEIN M	1.20e+00
41	51	73.9	281	1	YL82_CAEEL HYPOTHETICAL HISTONE 3	1.20e+00
42	51	73.9	288	1	YMH3_CAEEL HYPOTHETICAL HISTONE 3	1.20e+00
43	51	73.9	535	1	YGIS_ECOLI PUTATIVE BINDING PROTE	1.20e+00
44	51	73.9	3341	1	POLG_MCPFA GENOME POLYPROTEIN [CO	1.20e+00
45	50	72.5	505	1	YML8_YEAST HYPOTHETICAL 57.7 KD P	2.06e+00

ALIGNMENTS

RESULT 1  
ID H38\_STYLE STANDARD; PRT; 113 AA.  
AC P81202;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 15-DEC-1999 (Rel. 39, Last annotation update)  
DE HISTONE H3-8 (FRAGMENT).  
GN H3-8.  
OS Stylyonchia lemnae.  
OC Eukaryota; Alveolata; Ciliophora; hypotrichs; Stichotrichida;  
OC Oxytrichidae; Stylyonchia.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA BERNHARD D.;  
RT "Several highly divergent histone H3 genes in the hypotrich ciliate Stylyonchia lemnae".  
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: HISTONE H3, ALONG WITH HISTONE H4, PLAYS A CENTRAL ROLE IN NUCLEOSOME FORMATION.  
CC -!- SUBUNIT: THE NUCLEOSOME IS AN OCTAMER CONTAINING TWO MOLECULES OF H2A, H2B, H3, AND H4; WHICH WRAP APPROXIMATIVELY 146 BP OF DNA.  
CC -!- SIMILARITY: BELONGS TO THE HISTONE H3 FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; Y16634; CAA76337.1;  
CC PROSITE; PS00322; HISTONE\_H3\_1; FALSE\_NEG.  
CC PROSITE; PS00959; HISTONE\_H3\_2; 1.  
CC PFAM; PF00125; histone; 1.  
CC KW Nuclear protein; Chromosomal protein; DNA-binding; Nucleosome core.  
CC FT NON\_TER 1  
CC FT NON\_TER 113  
CC SEQUENCE 113 AA; 12685 MW; CA4374A6 CRC32;

Query Watch 100.0%; Score 69; DB 1; Length 113;  
Best Local Similarity 100.0%; Pred. No. 2.46e-05;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Db 32 RYRPGTVAL 40  
QY 1 RYRPGTVAL 9

PAGE BLANK (USPTO)

FEATURE  
9,27

#modified\_site N6-methyllysine, N6,N6-dimethyllysine or  
N6,N6,N6-trimethyllysine (Lys) (partial) #status  
experimental

SUMMARY #length 135 #molecular-weight 15257 #checksum 8291

Query Match 100.0%; Score 69; DB 1; Length 135;  
Best Local Similarity 100.0%; Pred. NO. 2.57e-04;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 40 RYREGTVAL 48

|||||||

Qy 1 RYREGTVAL 9

Search completed: Sat Apr 15 00:01:32 2000  
Job time : 9 secs.

#gene ch3-II  
#map\_position CH-II  
GENETICS NB2  
#gene ch3-III  
#map\_position CH-III  
CLASSIFICATION #superfamily histone H3  
KEYWORDS #superfamily histone H3  
SUMMARY #length 135 #molecular-weight 15310 #checksum 7476

Query Match 100.0%; Score 69; DB 2; Length 135;  
Best Local Similarity 100.0%; Pred. No. 2.57e-04;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 40 RYRPGTVAL 48  
QY 1 RYRPGTVAL 9  
|||||

RESULT 12  
ENTRY #type complete  
TITLE H3.3 like histone MH321 - mouse  
ORGANISM #formal\_name Mus musculus #common\_name house mouse  
DATE 24-Feb-1994 #sequence\_revision 24-Feb-1994 #text\_change 17-Mar-1999

ACCESSIONS JQ1984  
REFERENCE Wellman, S.E.; Casano, P.J.; Pilch, D.R.; Marzluff, W.F.; Sittman, D.B.  
#authors  
#journal Gene (1987) 59:29-39  
#title Characterization of mouse H3.3-like histone genes.  
#cross-references MUID:88137943  
#accession JQ1984  
#molecule\_type DNA  
#residues 1-135 #label WEL  
#note cross-reference

GENETICS  
#gene MH321  
CLASSIFICATION #superfamily histone H3  
KEYWORDS chromosomal protein  
SUMMARY #length 135 #molecular-weight 15224 #checksum 7489

Query Match 100.0%; Score 69; DB 2; Length 135;  
Best Local Similarity 100.0%; Pred. No. 2.57e-04;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 40 RYRPGTVAL 48  
QY 1 RYRPGTVAL 9  
|||||

RESULT 13  
ENTRY #type complete  
TITLE histone H3, embryonic - sea urchin (Strongylocentrotus purpuratus)  
ORGANISM #formal\_name Strongylocentrotus purpuratus #common\_name purple urchin  
DATE 30-Jun-1987 #sequence\_revision 30-Jun-1987 #text\_change 20-Mar-1998

ACCESSIONS A02629  
REFERENCE A93627  
#authors Kaemeyer, J.F.; Weinberg, E.S.  
#journal Nucleic Acids Res. (1986) 14:4557-4576  
#title Sequence, organization and expression of late embryonic H3 and H4 histone genes from the sea urchin, Strongylocentrotus purpuratus.

#cross-references MUID:86232591  
#accession A02629  
#molecule\_type DNA  
#residues 1-135 #label KAU  
#note #cross-references GB:X03952; NID:q10258; PID:q10258  
#note the authors translated the codon AGC for residue 96 as Arg and CGT for residue 102 as Gly

COMMENT This histone is expressed during late embryonic development.

CLASSIFICATION #superfamily histone H3  
KEYWORDS chromosomal protein; DNA binding; embryo; nucleosome core  
SUMMARY #length 135 #molecular-weight 15370 #checksum 8810

Query Match 100.0%; Score 69; DB 1; Length 135;  
Best Local Similarity 100.0%; Pred. No. 2.57e-04;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 40 RYRPGTVAL 48  
QY 1 RYRPGTVAL 9  
|||||

RESULT 14  
ENTRY #type complete  
TITLE histone H3 - Volvox carteri  
ORGANISM #formal\_name Volvox carteri  
DATE 30-Jun-1989 #sequence\_revision 30-Jun-1989 #text\_change 08-Sep-1997

ACCESSIONS S00940  
REFERENCE Mueller, K.; Schmitt, R.  
#authors  
#journal Nucleic Acids Res. (1988) 16:4121-4136  
#title Histone genes of Volvox carteri: DNA sequence and organization of two H3-H4 gene loci.  
#cross-references MUID:88234003  
#accession S00940  
#molecule\_type DNA  
#residues 1-135 #label MUE  
#note #cross-references EMBL:X08963; NID:q21983; PID:q21985

GENETICS  
#introns 46/1  
CLASSIFICATION #superfamily histone H3  
KEYWORDS chromosomal protein; DNA binding; nucleosome core; nucleus  
SUMMARY #length 135 #molecular-weight 15310 #checksum 7476

Query Match 100.0%; Score 69; DB 2; Length 135;  
Best Local Similarity 100.0%; Pred. No. 2.57e-04;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 40 RYRPGTVAL 48  
QY 1 RYRPGTVAL 9  
|||||

RESULT 15  
ENTRY #type complete  
TITLE histone H3 - smallmouth buffalo fish  
ORGANISM #formal\_name Ictiobus bubalus #common\_name smallmouth buffalo fish  
DATE 30-Jun-1987 #sequence\_revision 30-Jun-1987 #text\_change 16-Feb-1997

ACCESSIONS A02627  
REFERENCE Hooper, J.A.; Smith, E.L.; Sommer, K.R.; Chalkley, R.  
#authors J. Biol. Chem. (1973) 248:3275-3279  
#journal Histone III. IV. Amino acid sequence of histone III of the testes of the carp, Letiobus bubalus.  
#cross-references MUID:73166575  
#accession A02627  
#molecule\_type protein  
#residues 1-135 #label HOO  
#note Lys-9 is epsilon-N-monomethyllysine, epsilon-N-dimethyllysine, epsilon-N-trimethyllysine, or unmodified in 10, 13, 11, and 65% of the molecules, respectively  
Lys-27 is epsilon-N-monomethyllysine, epsilon-N-dimethyllysine, epsilon-N-trimethyllysine, or unmodified in 36, 40, 16, and 8% of the molecules, respectively

CLASSIFICATION #superfamily histone H3  
KEYWORDS chromosomal protein; DNA binding; methylated amino acid; nucleosome core

#submission submitted to the EMBL Data Library, February 1994  
 #description Cloning of two differentially expressed reverse transcription fragments of the histone 3 gene of leptochox acervorum (Hymenoptera Formicidae).  
 #accession S42065

##molecule\_type mRNA  
 ##residues 1-83 ##label BAU  
 ##cross-references EMBL:X77742; NID:g456194; PID:g456195

## GENETICS

#gene H3.1  
 #superfamily histone H3  
 #chromosomal protein; DNA binding; nucleosome core; nucleus  
 #length 83 #checksum 2931

## CLASSIFICATION

Query Match 100.0%; Score 69; DB 2; Length 83;  
 Best Local Similarity 100.0%; Pred. No. 2.57e-04;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 8 RYRPGTVAL 16

QY 1 RYRPGTVAL 9

## RESULT 8

ENTRY S04521 #type fragment  
 TITLE histone H3 (clone pH3c-1) - alfalfa (fragment)  
 ORGANISM #formal\_name Medicago sativa #common\_name alfalfa  
 DATE 28-Feb-1990 #sequence\_revision 28-Feb-1990 #text\_change 08-Sep-1997

## ACCESSIONS

REFERENCE S04521  
 #authors Wu, S.C.; Gyoergye, J.; Dudits, D.  
 #journal Nucleic Acids Res. (1989) 17:3057-3063  
 #title Polyadenylated H3 histone transcripts and H3 histone variants in alfalfa.

#cross-references MUID:89263717

#accession S04521

##molecule\_type mRNA

##residues 1-119 ##label WUS

##cross-references EMBL:X13676; NID:g19612; PID:g829279

CLASSIFICATION #superfamily histone H3

KEYWORDS DNA binding; nucleus

SUMMARY #length 119 #checksum 3419

Query Match 100.0%; Score 69; DB 2; Length 119;

Best Local Similarity 100.0%; Pred. No. 2.57e-04;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 24 RYRPGTVAL 32

QY 1 RYRPGTVAL 9

## RESULT 9

ENTRY A02630 #type fragments  
 TITLE histone H3 - fruit fly (Drosophila melanogaster) (fragments)  
 ORGANISM #formal\_name Drosophila melanogaster  
 DATE 31-Mar-1991 #sequence\_revision 31-Mar-1991 #text\_change 23-Feb-1997

## ACCESSIONS

REFERENCE A02630  
 #authors Goldberg, M.L.  
 #citation Ph.D. thesis, Stanford Univ., 1979  
 #accession A02630

##molecule\_type DNA

##residues 1-121 ##label GOL

##note the author translated the codon CCC for residue 31 as Ala

## GENETICS

#gene FlyBase:His3

##cross-references FlyBase:FBgn0001199

CLASSIFICATION #superfamily histone H3

KEYWORDS chromosomal protein; DNA binding; nucleosome core; nucleus

## SUMMARY #length 121 #checksum 3455

Query Match 100.0%; Score 69; DB 2; Length 121;

Best Local Similarity 100.0%; Pred. No. 2.57e-04;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 40 RYRPGTVAL 48

QY 1 RYRPGTVAL 9

## RESULT 10

ENTRY S59123 #type complete  
 TITLE histone H3 - Chlamydomonas reinhardtii  
 ORGANISM #formal\_name Chlamydomonas reinhardtii  
 DATE 15-Feb-1996 #sequence\_revision 01-Mar-1996 #text\_change 08-Sep-1997

## ACCESSIONS

REFERENCE S59123  
 #authors Walther, Z.; Hall, J.L.  
 #journal Nucleic Acids Res. (1995) 23:3756-3763  
 #title The uni chromosome of Chlamydomonas: histone genes and nucleosome structure.

#cross-references MUID:96017782

#accession S59123

##status preliminary

##molecule\_type DNA

##residues 1-135 ##label WAL

##cross-references EMBL:L41841; NID:g790699; PID:g790700

CLASSIFICATION #superfamily histone H3

SUMMARY #length 135 #molecular-weight 15260 #checksum 8478

## Query Match

100.0%; Score 69; DB 2; Length 135;

Best Local Similarity 100.0%; Pred. No. 2.57e-04;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 40 RYRPGTVAL 48

QY 1 RYRPGTVAL 9

## RESULT 11

ENTRY S59581 #type complete  
 TITLE histone H3 (clones CH-II and CH-III) - Chlamydomonas reinhardtii  
 ORGANISM #formal\_name Chlamydomonas reinhardtii  
 DATE 15-Feb-1996 #sequence\_revision 01-Mar-1996 #text\_change 13-Mar-1998

## ACCESSIONS

REFERENCE S59581; S59585  
 #authors Fabry, S.; Mueller, K.; Lindauer, A.; Park, P.B.; Cornelius, T.; Schmitt, R.  
 #journal Curr. Genet. (1995) 28:333-345  
 #title The organization structure and regulatory elements of Chlamydomonas histone genes reveal features linking plant and animal genes.

#accession S59581

##status nucleic acid sequence not shown

##molecule\_type DNA

##residues 1-135 ##label FAB

##cross-references EMBL:U16724; NID:g571469; PID:g571470

##experimental\_source clone CH-II

##genetics NB1

##note the authors did not translate the codon for residue 1

## #accession

S59585 nucleic acid sequence not shown

##status nucleic acid sequence not shown

##molecule\_type DNA

##residues 1-135 ##label FAW

##cross-references EMBL:U16725; NID:g571474; PID:g571475

##experimental\_source clone CH-III

##genetics NB2

##note the authors did not translate the codon for residue 1

GENETICS NB1



```
CLASSIFICATION #superfamily histone H3
SUMMARY #length 48 #checksum 171

Query Match 100.0%; Score 69; DB 2; Length 48;
Best Local Similarity 100.0%; Pred. No. 2.57e-04;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 40 RYRPGTVAL 48
QY 1 RYRPGTVAL 9

RESULT 3
ENTRY C61286 #type fragment
TITLE histone H3 - cycad (Encephalartos caffer) (fragment)
ORGANISM #formal_name Encephalartos caffer #common_name cycad
DATE 19-Mar-1997 #sequence_revision 19-Dec-1997 #text_change
26-Feb-1998

ACCESSIONS C61286
REFERENCE A61286 #type fragment
#authors Brandt, W.F.; Strickland, W.N.; Morgan, M.; Von Holt, C.
#journal FEBS Lett. (1974) 40:167-172
#title Comparison of the N-terminal amino acid sequences of histone
F3 from a mammal, a bird, a shark, an echinoderm, a mollusc
and a plant.
#accession C61286
#status preliminary
#molecule_type protein
#residues 1-48 #label BRA
CLASSIFICATION #superfamily histone H3
SUMMARY #length 48 #checksum 171

Query Match 100.0%; Score 69; DB 2; Length 48;
Best Local Similarity 100.0%; Pred. No. 2.57e-04;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 40 RYRPGTVAL 48
QY 1 RYRPGTVAL 9

RESULT 4
ENTRY S51664 #type fragment
TITLE histone H3.3 - tomato (fragment)
ORGANISM #formal_name Lycopersicon esculentum #common_name tomato
DATE 07-May-1995 #sequence_revision 01-Sep-1995 #text_change
24-Jul-1998

ACCESSIONS S51664
REFERENCE S51664
#authors Hartung, F.
#submission submitted to the EMBL Data Library, December 1994
#accession S51664
#molecule_type mRNA
#residues 1-60 #label HAR
#cross-references EMBL:X83422
#experimental_source cultivar Rutgers; clone Y17-3
CLASSIFICATION #superfamily histone H3
KEYWORDS DNA binding
SUMMARY #length 60 #checksum 8500

Query Match 100.0%; Score 69; DB 2; Length 60;
Best Local Similarity 100.0%; Pred. No. 2.57e-04;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 41 RYRPGTVAL 49
QY 1 RYRPGTVAL 9

RESULT 5
ENTRY B38309 #type fragments
TITLE histone H3.2 - alfalfa (fragments)
ORGANISM #formal_name Medicago sativa #common_name alfalfa

14-Jun-1991 #sequence_revision 14-Jun-1991 #text_change
23-Feb-1997
ACCESSIONS B38309
REFERENCE A38309
#authors Waterborg, J.H.
#journal J. Biol. Chem. (1990) 265:17157-17161
#title Sequence analysis of acetylation and methylation in two
histone H3 variants of alfalfa.
#cross-references MUID:91009145
#accession B38309
#status preliminary
#molecule_type protein
#residues 1-62 #label WAT
CLASSIFICATION #superfamily histone H3
KEYWORDS chromosomal protein; DNA binding; nucleosome core; nucleus
SUMMARY #length 62 #checksum 7971

Query Match 100.0%; Score 69; DB 2; Length 62;
Best Local Similarity 100.0%; Pred. No. 2.57e-04;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 40 RYRPGTVAL 48
QY 1 RYRPGTVAL 9

RESULT 6
ENTRY S42066 #type fragment
TITLE histone H3.3 - Leptothorax acervorum (fragment)
ORGANISM #formal_name Leptothorax acervorum
DATE 19-Mar-1997 #sequence_revision 18-Jul-1997 #text_change
05-Jun-1998

ACCESSIONS S42066; S42067
REFERENCE S42065
#authors Baur, A.; Stetzer, N.E.; Buschinger, A.; Zimmermann, F.K.
#submission submitted to the EMBL Data Library, February 1994
#description Cloning of two differentially expressed reverse transcription
fragments of the histone 3 gene of Leptothorax acervorum
(Hymenoptera Formicidae).
#accession S42066
#molecule_type mRNA
#residues 1-83 #label BAW
#cross-references EMBL:X77741; NID:g456196; PID:g456197
#experimental_source freshly laid eggs
#accession S42067
#molecule_type mRNA
#residues 1-83 #label BAW
#cross-references EMBL:X77740; NID:g459251; PID:g456199
#experimental_source larvae, pupae and adults
GENETICS H3.3
#gene #superfamily histone H3
CLASSIFICATION chromosomal protein; DNA binding; nucleosome core; nucleus
KEYWORDS #length 83 #checksum 1158
SUMMARY

Query Match 100.0%; Score 69; DB 2; Length 83;
Best Local Similarity 100.0%; Pred. No. 2.57e-04;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 8 RYRPGTVAL 16
QY 1 RYRPGTVAL 9

RESULT 7
ENTRY S42065 #type fragment
TITLE histone H3.1 - Leptothorax acervorum (fragment)
ORGANISM #formal_name Leptothorax acervorum
DATE 19-Mar-1997 #sequence_revision 18-Jul-1997 #text_change
08-Sep-1997

ACCESSIONS S42065
REFERENCE S42065
#authors Baur, A.; Stetzer, N.E.; Buschinger, A.; Zimmermann, F.K.
```

\*\*\*\*\*  
M P E R L H  
\*\*\*\*\*  
(TM)  
\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd  
MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Sat Apr 15 00:01:23 2000; Maspar time 3.21 Seconds  
Tabular output not generated. 112.216 Million cell updates/sec

Title: >US-08-452-843-11  
Description: (1-9) from US08452843.pep  
Perfect Score: 69  
Sequence: 1 RYRPGTVAL 9  
Scoring table: PAM 150  
Gap 15  
Searched: 122810 seqs, 40068593 residues  
Post-processing: Minimum Match 0%  
Listing first 45 summaries  
Database: p1r62  
1.p1r1 2.p1r2 3.p1r3 4.p1r4  
Statistics: Mean 23.546; Variance 28.097; scale 0.838

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.  
SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	69	100.0	48	2 B61286	histone H3 - sandpape	2.57e-04
2	69	100.0	48	2 A61286	histone H3 - sea urch	2.57e-04
3	69	100.0	48	2 C61286	histone H3 - cycad (E	2.57e-04
4	69	100.0	60	2 S1664	histone H3.2 - alfalf	2.57e-04
5	69	100.0	62	2 S38309	histone H3.2 - alfalf	2.57e-04
6	69	100.0	83	2 S42066	histone H3.3 - Leptot	2.57e-04
7	69	100.0	83	2 S42065	histone H3.1 - Leptot	2.57e-04
8	69	100.0	119	2 S04521	histone H3 (clone PH3	2.57e-04
9	69	100.0	121	2 A02630	histone H3 - fruit fl	2.57e-04
10	69	100.0	135	2 S59123	histone H3 - Chlamydo	2.57e-04
11	69	100.0	135	2 S59581	histone H3 (clones CH	2.57e-04
12	69	100.0	135	2 J01984	H3.3 like histone MH3	2.57e-04
13	69	100.0	135	1 HSUR3P	histone H3, embryonic	2.57e-04
14	69	100.0	135	2 S00940	histone H3 - Volvox c	2.57e-04
15	69	100.0	135	1 HSRI3	histone H3 - smallmou	2.57e-04
16	69	100.0	135	1 HSXL32	histone H3.2 - Africa	2.57e-04
17	69	100.0	136	2 I50244	histone 3.3A - chicke	2.57e-04
18	69	100.0	136	2 JH0304	histone H3.2 - mouse	2.57e-04
19	69	100.0	136	1 HSHU33	histone H3.3 - human	2.57e-04
20	69	100.0	136	1 HSKW3	histone H3 - Caenorha	2.57e-04
21	69	100.0	136	1 HSHU3	histone H3.1 - human	2.57e-04
22	69	100.0	136	1 HSZP3	histone H3.1 - fissio	2.57e-04
23	69	100.0	136	2 S01198	histone H3 - starfish	2.57e-04

24	69	100.0	136	2 S10097	histone H3 - fruit fl	2.57e-04
25	69	100.0	136	2 S01197	histone H3 - starfish	2.57e-04
26	69	100.0	136	2 S32638	histone H3.1 - Africa	2.57e-04
27	69	100.0	136	2 S06743	histone H3 - mouse	2.57e-04
28	69	100.0	136	2 I49395	histone H3.2 protein	2.57e-04
29	69	100.0	136	2 I49398	histone H3.1 protein	2.57e-04
30	69	100.0	136	2 S57473	histone H3 - human	2.57e-04
31	69	100.0	136	2 S10168	histone H3.3A - rabb1	2.57e-04
32	69	100.0	136	2 JQ0757	histone H3 - staghorn	2.57e-04
33	69	100.0	136	2 S01196	histone H3 - starfish	2.57e-04
34	69	100.0	136	2 S20678	histone H3 - starfish	2.57e-04
35	69	100.0	136	2 S24346	histone H3.3-like pro	2.57e-04
36	69	100.0	136	2 A56854	histone H3 - Tigrilopu	2.57e-04
37	69	100.0	136	2 S61218	histone H3.3 - fruit	2.57e-04
38	69	100.0	136	2 A45941	histone H3 - Atlantic	2.57e-04
39	69	100.0	136	2 S20669	histone H3 - starfish	2.57e-04
40	69	100.0	136	2 S34185	histone H3 - rat	2.57e-04
41	69	100.0	136	2 I57019	H3 histone - rat	2.57e-04
42	69	100.0	136	2 S61220	histone H3.3 - fruit	2.57e-04
43	69	100.0	136	2 S50140	H3.3 histone - sea ur	2.57e-04
44	69	100.0	136	2 I50460	H3 histone - muscovy	2.57e-04
45	69	100.0	136	2 A56580	histone H3 - midge (C	2.57e-04

ALIGNMENTS

RESULT 1  
ENTRY B61286 #type fragment  
TITLE histone H3 - Sandpape Limpet (fragment)  
ORGANISM histone H3 - Sandpape Limpet (fragment)  
DATE 12-May-1994 #sequence-revision 12-May-1994 #text\_change 03-May-1996  
ACCESSIONS B61286  
REFERENCE A61286  
#authors Brandt, W.F.; Strickland, W.N.; Morgan, M.; Von Holt, C.  
#journal FEBS Lett. (1974) 40:167-172  
#title Comparison of the N-terminal amino acid sequences of histone F3 from a mammal, a bird, a shark, an echinoderm, a mollusc and a plant.  
#accession B61286  
#status preliminary  
#molecule\_type protein  
#residues 1-48 #label BRA  
CLASSIFICATION #superfamily histone H3  
SUMMARY #length 48 #checksum 171

Query Match 100.0%; Score 69; DB 2; Length 48;  
Best Local Similarity 100.0%; Pred. No. 2.57e-04;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db	40 RYRPGTVAL 48	
Qy	1 RYRPGTVAL 9	
RESULT 2		
ENTRY A61286 #type fragment		
TITLE histone H3 - sea urchin (Parechinus angulosus) (fragment)		
ORGANISM histone H3 - sea urchin (Parechinus angulosus) (fragment)		
DATE 12-May-1994 #sequence-revision 12-May-1994 #text_change 03-May-1996		
ACCESSIONS A61286		
REFERENCE A61286		
#authors Brandt, W.F.; Strickland, W.N.; Morgan, M.; Von Holt, C.		
#journal FEBS Lett. (1974) 40:167-172		
#title Comparison of the N-terminal amino acid sequences of histone F3 from a mammal, a bird, a shark, an echinoderm, a mollusc and a plant.		
#accession A61286		
#status preliminary		
#molecule_type protein		
#residues 1-48 #label BRA		

THIS PAGE BLANK (USPTO)

CC induce an immunological response in a mammal or to identify inhibitors,  
 CC activators or novel antivirals. Antagonists of the proteins can be used  
 CC to inhibit a viral polypeptide. The DNA sequence or a vector containing  
 CC it can also be used to induce an immunological response in a mammal.  
 SQ Sequence 610 AA;

Query Match 66.7%; Score 46; DB 1; Length 610;  
 Best Local Similarity 71.4%; Pred. No. 1.52e+02;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 11 KYRPGTL 17  
 :||||:  
 Qy 1 RYRPGTV 7

RESULT 14  
 ID W2097 standard; Protein; 649 AA.

AC W2097;  
 DT 18-DEC-1998 (first entry)  
 DE HSV-2 strain SB5; Immunological response induction;  
 DE HSV-2 strain SB5; immunological response induction; therapy;  
 DE HSV-2 strain SB5; immunological response induction; therapy;  
 KW antiviral identification; viral protein inhibitor;  
 OS Herpes simplex virus type 2.  
 FN W09820018-A1.  
 PD 14-MAY-1998.  
 PF 31-OCT-1997; U20016.  
 PR 09-JUN-1997; US-049018.  
 PR 04-NOV-1996; US-030279.  
 PA (SMIK ) SMITHKLINE BEECHAM CORP.  
 PI Chan JY, Dabrowski-Amaral CE, Delvecchio AM, Dillon SB,  
 PI Esser KM, Leary JJ;  
 DR WPI; 98-286847/25.  
 DR N-PSDB; V62154.  
 PT Herpes simplex virus type-2 sequences - useful in, e.g. prevention  
 PT and treatment of infection or inducing immunological response in  
 PT mammal  
 PS Claim 10; Page 79-80; 748pp; English.  
 CC This sequence represents a Herpes simplex virus type-2 (HSV-2) protein  
 CC sequence of the invention. This sequence was isolated from a HSV-2 strain  
 CC SB5 (deposited as ATCC VR-2546) DNA fragment designated Contig ID 10.  
 CC The proteins can be used for the treatment or prevention of disease, to  
 CC induce an immunological response in a mammal or to identify inhibitors,  
 CC activators or novel antivirals. Antagonists of the proteins can be used  
 CC to inhibit a viral polypeptide. The DNA sequence or a vector containing  
 CC it can also be used to induce an immunological response in a mammal.  
 SQ Sequence 649 AA;

Query Match 66.7%; Score 46; DB 1; Length 649;  
 Best Local Similarity 71.4%; Pred. No. 1.52e+02;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 50 KYRPGTL 56  
 :||||:  
 Qy 1 RYRPGTV 7

RESULT 15  
 ID R97245 standard; Protein; 4472 AA.

AC R97245;  
 DT 07-JAN-1997 (first entry)  
 DE Virulence gene cluster polypeptide product.  
 DE Mutant; adaptation; virulence factor; identification; screening;  
 KW vaccine; drugs; infection; treatment.  
 OS Salmonella typhimurium.  
 EH Key Location/Qualifiers  
 FT region  
 FT /note= "All x's in this sequence correspond to  
 FT termination codons in the virulence gene  
 FT cluster sequence given in T09224."  
 PN W09617951-A2.  
 PD 13-JUN-1996.  
 PF 11-DEC-1995; G02875.  
 PR 09-DEC-1994; GB-024921.

PR 31-JAN-1995; GB-001881.  
 PR 05-MAY-1995; GB-009239.  
 PA (RPMs-) RPMs TECHNOLOGY LTD.  
 PI Holden DM;  
 DR WPI; 96-287194/29.  
 PT Identifying virulence genes in microorganisms - by introducing  
 PT mutants with insertion inactivated genes into environment and  
 PT retrieval and analysis of mutants  
 PS Claim 51; Figure 11; 131pp; English.  
 CC A method for identifying a microorganism having a reduced adaptation  
 CC to a particular environment comprising the steps of: (1) providing a  
 CC plurality of microorganisms each of which is independently mutated by  
 CC the insertional inactivation of a gene with a nucleic acid comprising  
 CC a unique marker sequence so that each mutant contains a different  
 CC marker sequence, or clones of the said microorganism; (2) providing  
 CC individually a stored sample of each mutant produced by step (1) and  
 CC providing individually stored nucleic acid comprising the unique  
 CC marker sequence from each individual mutant; (3) introducing a  
 CC plurality of mutants produced by step (1) into the said particular  
 CC environment and allowing those microorganisms which are able to do so  
 CC to grow in the said environment; (4) retrieving microorganisms from  
 CC the said environment or a selected part thereof and isolating the  
 CC nucleic acid from the retrieved microorganisms; (5) comparing any  
 CC marker sequences in the nucleic acid isolated in step (4) to the  
 CC unique marker sequence of each individual mutant stored as in step  
 CC (2); and (6) selecting an individual mutant which does not contain any  
 CC of the marker sequences as isolated in step (4). The products and  
 CC methods can be used for identifying virulence genes in microorganisms.  
 CC The mutant microorganisms can be used in vaccines or to screen for  
 CC drugs which reduce virulence or compounds useful for preventing,  
 CC ameliorating or treating infections in animals or plants.  
 SQ Sequence 4472 AA;

Query Match 66.7%; Score 46; DB 1; Length 4472;  
 Best Local Similarity 62.5%; Pred. No. 1.52e+02;  
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 627 YRTGSVGL 634  
 :|||l:l|  
 Qy 2 YRPGTV 9

Search completed: Sat Apr 15 00:01:05 2000  
 Job time : 42 secs.

ID W06491 standard; Protein; 332 AA.  
AC W06491;  
DE 05-FEB-1997 (first entry)  
DT Beta-1,4-galactosyltransferase-related protein #2.  
KW Murine; beta-1,4-galactosyltransferase-related protein; sterility;  
KW fertilisation; F9 cancer cell; Huynh's method.  
OS Mus musculus.  
PN J08196279-A.  
PD 06-AUG-1996.  
PF 25-JAN-1995; 009642.  
PR 25-JAN-1995; JP-009642.  
PA (MITK ) MITSUI TOATSU CHEM INC.  
PA (MURA) MURAMATSU T.  
DR WPI; 96-406013/41.  
DR N-PSDB; T45082.  
PT DNA sequence encoding beta-1,4-galactosyltransferase-related  
PT protein - useful for sterility diagnosis, and for assisting or  
PT inhibiting fertilisation  
PS Claim 4; Page 7-9; 11pp; Japanese.  
CC The sequences given in W06490-91 represent two clones of murine  
CC beta-1,4-galactosyltransferase-related proteins. These proteins  
CC can be used as diagnostic agents for various diseases. They are  
CC esp. useful in the diagnosis of sterility and in the aiding and  
CC inhibiting of fertilisation. The cDNA's encoding the two beta-1,4-  
CC galactosyltransferase-related proteins were isolated from F9 cancer  
CC cells according to Huynh's method.  
SQ Sequence 332 AA;

Query Match 66.7%; Score 46; DB 1; Length 332;  
Best Local Similarity 62.5%; Pred. No. 1.52e+02;  
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 39 FRPGVDL 46  
:|||||  
QY 2 YRPGTVAL 9

RESULT 11  
ID W06159 standard; Protein; 359 AA.  
AC W06159;  
DE 27-APR-1999 (first entry)  
DT Fructose-1,6-bisphosphate aldolase (FDA).  
DE Fructose-1,6-bisphosphate aldolase; FDA; carbon assimilation;  
KW starch; sucrose; crops; yield; growth; transgenic plants; potato;  
KW photosynthesis.  
KW Escherichia coli.  
OS Escherichia coli.  
PN W0988069-A1.  
PD 23-DEC-1998.  
PF 16-JUN-1998; U12447.  
PR 17-JUN-1997; US-049955.  
PA (MONS ) MONGANTO CO.  
PI Barry GF, Cheikh N, Kishore GM;  
DR WPI; 99-095343/08.  
DR N-PSDB; X08920.  
PT Use of fructose-1,6-triphosphate aldolase DNA - useful for, e.g.  
PT producing transgenic plants with increased photosynthesis rates,  
PT increased yields, increased growth rates and improved solids  
PT uniformity  
PS Example 1; Page 49-50; 75pp; English.  
CC Fructose-1,6-bisphosphate aldolase (FDA) catalyses the reversible  
CC reaction converting triosephosphate into fructose-1,6-bisphosphate  
CC aldolase. By inserting the FDA gene into expression vectors and  
CC inserting these vectors into the chloroplasts of plant cells,  
CC increased starch production can be achieved. Increasing the  
CC expression of the FDA enzyme in the chloroplast increases carbon  
CC assimilation and results in an increase in chloroplast starch  
CC production. This increase in carbon assimilation is a desirable  
CC trait in crop plants and leads to increased plant growth, storage  
CC ability, yield, vigour, and stress tolerance. Increasing FDA  
CC expression in the cytosol of photosynthetic cells leads to an  
CC increase in sucrose production. The transgenic plants containing  
CC the recombinant DNA can have increased photosynthesis rates,  
CC increased yields, increased growth rates and improved solids

CC uniformity compared with plants that do not contain the recombinant  
CC DNA molecule. Vectors containing the recombinant FDA gene are used  
CC particularly for improving potato products.

SQ Sequence 359 AA;

Query Match 66.7%; Score 46; DB 1; Length 359;  
Best Local Similarity 62.5%; Pred. No. 1.52e+02;  
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 230 YKPGNVVL 237  
:|||||  
QY 2 YRPGTVAL 9

## RESULT 12

ID R05272 standard; protein; 441 AA.  
AC R05272;  
DE 15-AUG-1990 (first entry)  
DT Polypeptide with amino peptidase-P activity encoded by new gene  
KW Amino peptidase-P.  
PN J02002373-A.  
PD 08-JAN-1990.  
PF 25-MAR-1989; 071138.  
PR 25-MAR-1989; JP-071138, JP-156193.  
PA (AJIN) Ajinomoto KK.  
PI WPI; 90-053424/08.  
DR N-PSDB; Q91838.  
PT Amino peptidase-P-coding gene -  
PT used in gene-provided recombinant DNA and recombinant  
PT DNA-provided survival cell stock  
PS Disclosure; 15pp; Japanese.  
CC It is new. Also new are recombinant DNA contg. its encoding DNA, cells  
CC transformed with the recombinant DNA, and prodn. of it by culturing the  
CC cells. The method allows economical, high yielding prodn. of it. It is  
CC also useful in separating or refining the enzyme.  
SQ Sequence 441 AA;

Query Match 66.7%; Score 46; DB 1; Length 441;  
Best Local Similarity 75.0%; Pred. No. 1.52e+02;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 305 YRPGTSIL 312  
:|||||  
QY 2 YRPGTVAL 9

## RESULT 13

ID W72228 standard; Protein; 610 AA.  
AC W72228;  
DE 13-JAN-1999 (first entry)  
DE HSV-2 strain SB5 Contig ID 15 ORF41a protein.  
KW HSV-2 strain SB5; immunological response induction; therapy;  
KW antiviral identification; viral protein inhibitor.  
OS Herpes simplex virus type 2.  
PN W09820016-A1.  
PD 14-MAY-1998.  
PF 31-OCT-1997; U20016.  
PR 09-JUN-1997; US-049018.  
PR 04-NOV-1996; US-030279.  
PA (SMIK ) SMITHKLINE BEECHAM CORP.  
PI Chan JY, Dabrowski-Amara CE, Delvecchio AM, Dillon SB,  
PI Esser KM, Leary JU;  
DR WPI; 98-286847/25.  
DR N-PSDB; V62176.  
PT Herpes simplex virus type-2 sequences - useful in, e.g. prevention  
PT and treatment of infection or inducing immunological response in  
PT mammal  
PS Claim 10; Page 143-144; 748pp; English.  
CC This sequence represents a Herpes simplex virus type-2 (HSV-2) protein  
CC sequence of the invention. This sequence was isolated from a HSV-2 strain  
CC SB5 (deposited as ATCC VR-2546) DNA fragment designated Contig ID 15.  
CC The proteins can be used for the treatment or prevention of disease, to

RESULT 6  
ID W75420 standard; protein; 1195 AA.  
AC W75420.  
DT 16-MAR-1999 (first entry)  
DE T.thermophilus nitrate reductase alpha subunit.  
KW Heat-stable; nitrate reductase; temperature; detection; food; toxicity;  
KW carcinogen.  
OS Thermus thermophilus.  
FH Key Location/Qualifiers  
FT Misc\_difference 630 /label= unknown  
FT Misc\_difference 669 /label= unknown  
FT Misc\_difference 691 /label= unknown  
FT ES2121561-A1.  
PN 16-NOV-1998.  
PD 09-MAY-1997; 001003.  
PF 09-MAY-1997; ES-001003.  
PR (UYNM-) UNIV AUTONOMA MADRID.  
PA WPI; 98-001909/01.  
DR Heat stable nitrate reductase for high temperature nitrate detection  
PT - comprises Thermus thermophilus derivative enhancing nitrite or  
PT nitrate reduction  
PS Disclosure; Fig 2; 8pp; Spanish.  
CC This sequence represents the amino acid sequence of the Thermus  
CC thermophilus heat-stable nitrate reductase alpha subunit. Heat stable  
CC nitrate reductase can be used for high-temperature detection of nitrates  
CC in samples, e.g. in food, where high levels of nitrates can be toxic or  
CC carcinogenic.  
SQ Sequence 1195 AA;

Query Match 59.6%; Score 48; DB 1; Length 1195;  
Best Local Similarity 75.0%; Pred. No. 8.71e+01;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
Db 523 YRAGIVAL 530  
||| :|||  
Qy 2 YRPGTVAL 9

RESULT 7  
ID R74668 standard; peptide; 19 AA.  
AC R74668.  
DT 19-DEC-1995 (first entry)  
DE HLA-DRB1\*0405 binding oligopeptide (VI).  
KW Oligopeptide; HLA-DRB1\*0405; immunosuppressant; lymphocyte;  
KW Epstein-Barr virus; B cell line.  
OS Homo sapiens.  
PN J07082295-A.  
PD 28-MAR-1995.  
PF 13-SEP-1993; 227091.  
PR 13-SEP-1993; JP-227091.  
PA (TEIJ ) TEIJIN LTD.  
DR WPI; 95-158991/21.  
PT Oligopeptide immunosuppressant - isolated from B lymphocytes of  
PT HLA-DRB1-0405 subjects or prep. by peptide synthesis  
PS Claim 2; Page 2; 8pp; Japanese.  
CC The sequences given in R74663-68 represent oligopeptides which bind  
CC to HLA-DRB1\*0405. These peptides act as immunosuppressants and are  
CC administered at a daily dose of 1-100 mg/kg. These peptides may be  
CC derived from lymphocytes derived from a patient having HLA-DRB1\*0405  
CC and treated with Epstein-Barr virus to give a B cell line to produce  
CC the peptides.  
SQ Sequence 19 AA;  
Query Match 68.1%; Score 47; DB 1; Length 19;  
Best Local Similarity 75.0%; Pred. No. 1.15e+02;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
Db 6 YRPAVAL 13  
||| :|||

Qy 2 YRPGTVAL 9

RESULT 8  
ID R83693 standard; peptide; 19 AA.  
AC R83693.  
DT 10-APR-1996 (first entry)  
DE HLA binding peptide homologous to pyruvate kinase M2 isozyme.  
KW HLA binding oligopeptide; immunosuppressant; autoimmune disease;  
KW pyruvate kinase; M2 isozyme; residues 101-119; homologue.  
OS Synthetic.  
PN J07208896-A.  
PD 08-AUG-1995.  
PF 20-JAN-1994; 004615.  
PR 20-JAN-1994; JP-004615.  
PA (TEIJ ) TEIJIN LTD.  
DR WPI; 95-309097/40.  
PT New HLA binding oligopeptide(s) - useful as immunosuppressants for  
PT treating auto-immune diseases  
PS Example 1; Page 5; 9pp; Japanese.  
CC The present peptide is homologous to the pyruvate kinase M2 isozyme  
CC residues 101-119, and is a HLA binding oligopeptide. It can be used  
CC as an immunosuppressant for the treatment of autoimmune diseases.  
SQ Sequence 19 AA;

Query Match 68.1%; Score 47; DB 1; Length 19;  
Best Local Similarity 75.0%; Pred. No. 1.15e+02;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
Db 6 YRPAVAL 13  
||| :|||  
Qy 2 YRPGTVAL 9

RESULT 9  
ID R03457 standard; protein; 16 AA.  
AC R03457;  
DT 02-AUG-1990 (first entry)  
DE Intracellular retention moiety derived from Thymine 2-1.  
KW Intracellular retention moiety; Thymine 2-1; tumour therapy.  
OS Synthetic.  
PN EP-359347-A.  
PD 21-MAR-1990.  
PF 14-AUG-1989; 250014.  
PR 15-AUG-1988; US-232337.  
PA (NEOR-) Neorx Corp.  
PI Anderson DC, Morgan AC, Abrams PG, Nichols EJ, Fritzberg AR;  
DR WPI; 90-085154/12.  
PT Covalently linked complex for tumour treatment - comprises  
PT treating with protein, cytotoxic agent and enhancing moiety.  
PS Claim 16; Page 22; 23pp; English.  
CC The sequence is one of several possible intracellular retention moieties  
CC which can be covalently attached to one or more other enhancing moieties  
CC such as an internalization moiety, and to a targeting protein and a  
CC cytotoxic agent. The moiety is designed to bind noncovalently to dsDNA  
CC in the cell so increasing the amt. of time that the targeting protein  
CC conjugate is retained intracellularly. The N-terminal Cys and Gly  
CC residues are added to allow covalent cross linking to the targeting  
CC protein. The C-terminal is amidated. The complex is useful for treatment  
CC and diagnosis of tumours.  
SQ See also R03435-60.  
SQ Sequence 16 AA;

Query Match 66.7%; Score 46; DB 1; Length 16;  
Best Local Similarity 75.0%; Pred. No. 1.52e+02;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
Db 9 YRRTVA 16  
||| :|||  
Qy 1 YRPGTVA 8  
||| :|||  
RESULT 10

DE Histone H3.3 like protein.  
KW Human cDNA; library; enzyme; protein.  
OS Homo sapiens.  
PN WO9403599-A.  
PD 17-FEB-1994.  
PF 04-AUG-1993; J01095.  
PR 04-AUG-1992; JP-208077.  
PR 13-NOV-1992; JP-327619.  
PR 26-FEB-1993; JP-061431.  
PA (SAGA) SAGAMI CHEM RES CENTRE.  
PI Iwahori A, Kato S, Kato T, Kim N, Oh S, Sekine S;  
DR WPI: 94-065688/08.  
DR N-PSDB: Q57414.

PT cDNA of human origin and proteins coded by it - which may be expressed by in vivo or in vitro translation using sense RNA or PT antisense DNA corresponding to the cDNA.  
PS Claim 1; Page 29; 167pp; Japanese.  
CC mRNA expressed in human fibrosarcoma cell line HT-1080 was isolated and used to construct a cDNA library using vector CC pKAL. Clone HP00014 encoding histone H3.3-like protein CC was isolated.  
SQ Sequence 70 AA;

Query Match 100.0%; Score 69; DB 1; Length 70;  
Best Local Similarity 100.0%; Pred. No. 1.69e-01;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 41 RYRPGTVAL 49  
|||||  
QY 1 RYRPGTVAL 9

## RESULT 3

ID Y07284 standard; protein; 134 AA.  
AC Y07284;  
DT 06-JUL-1999 (first entry)  
DE Histone H3 consensus sequence.  
KW Consensus; histone H4; mammalian; amphibian; reptilian; transfection.  
OS Mammalia.  
OS Amphibia.  
OS Reptilia.  
PN EP-908521-A1.  
PD 14-APR-1999.  
PF 10-OCT-1997; 117574.  
PR 10-OCT-1997; EP-117574.  
PA (HMRI) HOECHST MARION ROUSSEL DEUT GMBH.  
PI Chandra A, Chandra P, Dermirhan I, Hasselmayr O;  
DR WPI: 99-217072/19.  
PT New transfection system comprising a histone protein, useful in gene therapy and drug screening assays  
PS Disclosure; Page 16; 32pp; English.  
CC This sequence represents a consensus sequence found across histone H4 proteins from mammalian, amphibian and reptilian animals. The invention CC relates to the use of a transfection system that comprises a histone CC protein or derivative, and a nucleic acid.  
SQ Sequence 134 AA;

Query Match 100.0%; Score 69; DB 1; Length 134;  
Best Local Similarity 100.0%; Pred. No. 1.69e-01;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 41 RYRPGTVAL 49  
|||||  
QY 1 RYRPGTVAL 9

## RESULT 4

ID Y07285 standard; protein; 135 AA.  
AC Y07285;  
DT 06-JUL-1999 (first entry)  
DE Human histone H3 sequence.  
KW Consensus; histone H4; human; transfection.  
OS Homo sapiens.

PN EP-908521-A1.  
PD 14-APR-1999.  
PF 10-OCT-1997; 117574.  
PR 10-OCT-1997; EP-117574.  
PA (HMRI) HOECHST MARION ROUSSEL DEUT GMBH.  
PI Chandra A, Chandra P, Dermirhan I, Hasselmayr O;  
DR WPI: 99-217072/19.  
PT New transfection system comprising a histone protein, useful in gene therapy and drug screening assays  
PS Disclosure; Page 16; 32pp; English.  
CC This sequence represents the human histone H3 protein. The invention CC relates to the use of a transfection system that comprises a histone CC protein or derivative, and a nucleic acid.  
SQ Sequence 135 AA;

Query Match 100.0%; Score 69; DB 1; Length 135;  
Best Local Similarity 100.0%; Pred. No. 1.69e-01;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 40 RYRPGTVAL 48  
|||||  
QY 1 RYRPGTVAL 9

## RESULT 5

ID W47029 standard; Protein; 566 AA.  
AC W47029;  
DT 06-JUL-1998 (first entry)  
DE Human N-proteinase (70 kDa short form).  
KW N-proteinase; human; collagen; antibody; rheumatoid arthritis;  
KW fibrosis; Ehlers-Danlos disease; diagnosis; therapy.  
OS Homo sapiens.  
PN WO9800555-A1.  
PD 08-JAN-1998.  
PF 03-JUL-1997; U12427.  
PR 02-JUL-1997; US-886333.  
PR 03-JUL-1996; US-021203.  
PA (COLI)/ COLIGE A.  
PA (LAPI)/ LAPIERE C.  
PA (PROC)/ PROCKOP D J.  
PI Collige A, Lapiere C, Prockop DJ;  
DR WPI: 98-086980/08.  
DR N-PSDB: V06593.

PT Polynucleotide sequence encoding human N-proteinase - used to produce mature collagen in vitro and antibodies to treat fibrosis and rheumatoid arthritis  
PS Disclosure; Fig 2B; 49pp; English.  
CC This polypeptide comprises the 70 kDa short form of human N-proteinase, the enzyme responsible for cleaving N-propeptide from procollagen to produce mature collagen. The amino acid sequence CC was deduced from a cDNA sequence (see V06593) derived from human skin fibroblast cDNA clones. The 130 kDa long form (see W47028) of CC human N-proteinase was also identified. Human N-proteinases CC can be produced using a claimed method in which a host cell is CC transformed or transfected with an N-proteinase polynucleotide CC sequence, cultured in an appropriate culture medium, and the CC N-proteinase is isolated from the medium. The N-proteinase can be CC used for the production of mature collagen in vitro and for the CC production of antibodies which may be used for diagnosis and CC therapy of diseases including fibrosis and rheumatoid arthritis. CC The N-proteinase may also be administered to treat a disease CC resulting from insufficient production of N-proteinase, such as CC Ehlers-Danlos disease.  
SQ Sequence 566 AA;

Query Match 69.6%; Score 48; DB 1; Length 566;  
Best Local Similarity 62.5%; Pred. No. 8.71e-01;  
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 543 KFRPGAVA 550  
:::|::|  
QY 1 RYRPGTVA 8

\*\*\*\*\*  
W P S R L  
\*\*\*\*\* (TM)  
\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Sat Apr 15 00:00:23 2000; Maspar time 6.60 Seconds  
Tabular output not generated. 32.295 Million cell updates/sec

Title: >US-08-452-843-11  
Description: (1-9) from US08452843.pap  
Perfect Score: 69  
Sequence: 1 RYRPGTVAL 9

Scoring table: PAM 150  
Gap 15

Searched: 188963 seqs, 23686106 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: a-geneseq36  
1:geneseqp

Statistics: Mean 17.319; Variance 44.133; scale 0.392

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	69	100.0	9	1 R9372	Histone H3.3 derived i	1.69e-01
2	69	100.0	70	1 R46075	Histone H3.3 like prot	1.69e-01
3	69	100.0	134	1 Y07284	Histone H3 consensus se	1.69e-01
4	69	100.0	135	1 Y07285	Human histone H3 seque	1.69e-01
5	48	69.6	566	1 W47029	Human N-proteinase (70	8.71e-01
6	48	69.6	1195	1 W75420	T.thermophilus nitrate	8.71e-01
7	47	68.1	19	1 R74668	HLA-DRB1*0405 binding	1.15e-02
8	47	68.1	19	1 R83693	HLA binding peptide bo	1.15e-02
9	46	66.7	16	1 R03457	Intracellular retentio	1.52e-02
10	46	66.7	332	1 W06491	Beta-1-4-galactosyltra	1.52e-02
11	46	66.7	359	1 W96159	Fructose-1,6-bisphosph	1.52e-02
12	46	66.7	441	1 R05272	Polypeptide with amino	1.52e-02
13	46	66.7	610	1 W22228	HSV-2 strain SB5 Conti	1.52e-02
14	46	66.7	649	1 W22097	HSV-2 strain SB5 Conti	1.52e-02
15	46	66.7	4472	1 R37245	Virulence gene cluster	1.52e-02
16	45	65.2	31	1 R21422	Matrix peptide from bo	1.99e-02
17	45	65.2	52	1 R12875	Non-collagenous bone m	1.99e-02
18	45	65.2	52	1 R12876	Non-collagenous bone m	1.99e-02
19	45	65.2	239	1 R99423	Mucin-derived protein	1.99e-02
20	45	65.2	240	1 R99422	Mucin-derived protein	1.99e-02
21	45	65.2	255	1 R99420	Mucin-derived protein	1.99e-02
22	45	65.2	264	1 R99421	Mucin-derived protein	1.99e-02
23	45	65.2	273	1 R99418	Mucin-derived protein	1.99e-02

24	45	65.2	282	1 R89419	Mucin-derived protein	1.99e-02
25	45	65.2	327	1 R96298	Glycoprotein 39 C term	1.99e-02
26	45	65.2	348	1 R27662	C-terminal region of H	1.99e-02
27	45	65.2	391	1 W52355	Synthetic lysyl oxidas	1.99e-02
28	45	65.2	417	1 W52356	Homo sapiens lysyl oxi	1.99e-02
29	45	65.2	455	1 R23973	Transmembrane form of	1.99e-02
30	45	65.2	497	1 R81462	Human derived cytochro	1.99e-02
31	45	65.2	497	1 R72375	Human auxillary cytoch	1.99e-02
32	45	65.2	497	1 R72377	Human auxillary cytoch	1.99e-02
33	45	65.2	497	1 W4869	Cytochrome P4501Id6	1.99e-02
34	45	65.2	497	1 R93185	Human cytochrome P450	1.99e-02
35	45	65.2	497	1 R72376	Human auxillary cytoch	1.99e-02
36	45	65.2	497	1 R93184	Human cytochrome P450	1.99e-02
37	45	65.2	497	1 R72378	Human auxillary cytoch	1.99e-02
38	45	65.2	497	1 R93183	Human cytochrome P450	1.99e-02
39	45	65.2	497	1 R93182	Human cytochrome P450	1.99e-02
40	45	65.2	523	1 R71976	Pertussis A.	1.99e-02
41	45	65.2	770	1 W34199	Streptomyces efflux pu	1.99e-02
42	45	65.2	770	1 W55800	Streptomyces roseofulv	1.99e-02
43	45	65.2	824	1 W23274	Bordetella pertussis p	1.99e-02
44	45	65.2	1528	1 W33363	Human multidrug resist	1.99e-02
45	45	65.2	3164	1 R94345	Hepatitis GB virus (HG	1.99e-02

ALIGNMENTS

RESULT 1  
ID R9372 standard; peptide; 9 AA.  
AC R9372; 1996 (first entry)  
DE Histone H3.3 derived immunogenic peptide.  
KW Immunogenic peptide; supermotif; HLA molecule; CTL response;  
KW therapeutic; diagnostic; cancer; viral infection; hepatitis B;  
KW hepatitis C.  
OS Synthetic C.  
PN WO9603140-A1.  
PD 08-FEB-1996.  
PF 21-JUL-1995; U09234.  
PR 21-JUL-1994; US-278634.  
PR 23-NOV-1994; US-344824.  
PR 30-MAY-1995; US-452843.  
PA (CITE-) CYTEL CORP.  
PI Sette A, Sidney J;  
DR WPI; 96-116784/12.  
PT Compsn. comprising immunogenic peptide with supermotif allowing more  
PT than one HLA mol. to bind - used to induce CTL response in patient  
PT and for in vivo and ex vivo therapeutic and diagnostic applications  
PS Claim 2; Page 26; 32pp; English.  
CC The sequences given in R9362-82 are immunogenic peptides which were  
CC use in the composition of the invention. The composition comprises  
CC an immunogenic peptide of 9-10 residues with a supermotif which  
CC allows binding of more than one HLA molecule. It pref. comprises  
CC two conserved residues, a first at the 2nd position from the N-  
CC terminal is Pro, and a 2nd at the C-terminal is Met. These peptides  
CC are used to induce a CTL response in a patient. They are also  
CC useful in compositions for in vivo and ex vivo therapeutic and  
CC diagnostic applications, e.g the treatment of cancer and viral  
CC infections, e.g. hepatitis B and C.  
SQ Sequence 9 AA;

Query Match 100.0%; Score 69; DB 1; Length 9;  
Best Local Similarity 100.0%; Pred. No. 1.69e-01;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db	1 RYRPGTVAL 9
Qy	1 RYRPGTVAL 9
RESULT 2	
ID R46075; 70 AA.	
DT 19-OCT-1994 (first entry)	



THIS PAGE BLANK (uspto)

Search completed: Fri Apr 14 23:56:37 2000  
Job time : 46 secs.

RL Nucleic Acids Res. 12:2917-2928(1984).  
RN [3]  
RX SEQUENCE FROM N.A. (P3).  
RX MEDLINE: 85182627.  
RA GONZALEZ F.J., KIMURA S., NEBERT D.W.;  
RT "Comparison of the flanking regions and introns of the mouse 2,3,7,8-  
RT tetrachlorodibenzo-p-dioxin-inducible cytochrome P1-450 and P3-450  
RT genes.";  
RL J. Biol. Chem. 260:5040-5049(1985).  
RN [4]  
RP ERRATUM.  
RA GONZALEZ F.J., KIMURA S., NEBERT D.W.;  
RL J. Biol. Chem. 260:11884-11889(1985).  
RN [5]  
RX SEQUENCE FROM N.A. (P3).  
RX MEDLINE: 85028449.  
RA GONZALEZ F.J., MACKENZIE P.I., KIMURA S., NEBERT D.W.;  
RT "Isolation and characterization of full-length mouse cDNA and genomic  
RT clones of 3-methylcholanthrene-inducible cytochrome P1-450 and  
RT P3-450.";  
RL Gene 29:281-292(1984).  
RN [6]  
RP SEQUENCE FROM N.A. (P2).  
RC STRAIN-DBA/2N; TISSUE=LIVER;  
RX MEDLINE: 86312932.  
RA KIMURA S., NEBERT D.W.;  
RT "cDNA and complete amino acid sequence of mouse P2(450): allelic  
RT variant of mouse P3(450) gene.";  
RL Nucleic Acids Res. 14:6765-6766(1986).  
CC -1- FUNCTION: CYTOCHROMES P450 ARE A GROUP OF HEME-THIOLATE  
CC MONOOXYGENASES. IN LIVER MICROSOMES, THIS ENZYME IS INVOLVED IN AN  
CC NADPH-DEPENDENT ELECTRON TRANSPORT PATHWAY. IT OXIDIZES A VARIETY  
CC OF STRUCTURALLY UNRELATED COMPOUNDS, INCLUDING STEROIDS, FATTY  
CC ACIDS, AND XENOBIOTICS.  
CC -1- CATALYTIC ACTIVITY: RH + REDUCED FLAVOPROTEIN + O(2) -> ROH +  
CC OXIDIZED FLAVOPROTEIN + H(2)O.  
CC -1- SUBCELLULAR LOCATION: MEMBRANE-BOUND. ENDOPLASMIC RETICULUM.  
CC -1- INDUCTION: BY 3-METHYLCOLANTHRENE (3MC).  
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; X01682; CAA25837.1; -;  
DR EMBL; X00479; CAA25156.1; -;  
DR EMBL; X04283; CAA27832.1; -;  
DR EMBL; K02589; AAA37509.1; ALT\_SEQ.  
DR EMBL; M10022; AAA37508.1; -;  
DR PIR; A00186; O4MSN3.  
DR PIR; B23923; B23923.  
DR MGD; MGI:88589; CYP1A2.  
DR PROSITE; PS00086; CYTOCHROME\_P450; 1.  
DR PEAM; PF00067; P450; 1.  
KW Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme;  
KW Microsome; Endoplasmic reticulum; Polymorphism.  
FT BINDING 456 HEME.  
FT VARIANT 384 384 I -> M (IN P2).  
SQ SEQUENCE 513 AA; 58184 MW; 9B334C3C CRC32;  
  
Query Match 70.1%; Score 54; DB 1; Length 513;  
Best Local Similarity 71.4%; Pred. No. 3.64e+00;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
  
Db 238 FVLYRL 244  
QY 3 FVLYRL 9

Query Match 71.4%; Score 55; DB 1; Length 462;  
Best Local Similarity 77.8%; Pred. No. 2.30e+00;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 269 IPWIVRYL 277  
QY 1 IPFPIVRYL 9

RESULT 14  
ID GIPR\_HUMAN STANDARD; PRT; 466 AA.  
AC P48546; Q16400; Q14401;  
DT 01-FEB-1996 (Rel. 33, Last sequence update)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)  
DE GASTRIC INHIBITORY POLYPEPTIDE RECEPTOR (GIP-R) (GLUCOSE-  
DEPENDENT INSULINOTROPIC POLYPEPTIDE RECEPTOR).  
DE GIPR.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Primates; Catarrhini; Homiidae; Homo.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA USLIN T.B., GRUBER C., MODI W., BONNER T.I.;  
RL Submitted (OCT-1995) to the EMBL/GenBank/DDBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA VOLZ A., GOKE R., LANKAT-BUTTEREIT B., FEHMANN H.C., BODE H.P.,  
RA "Molecular cloning, functional expression, and signal transduction of  
RT the GIP-receptor cloned from a human insulinoma.";  
RL FEBS Lett. 373:23-29(1995).  
RN [3]  
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.  
RC TISSUE-PANCREAS;  
RX MEDLINE: 96007224.  
RA GREMLICH S., FORRETT A., HANI E.H., CHERIF D., VIONNET N., FROGUEL P.,  
RA THORENS B.;  
RT "Cloning, functional expression, and chromosomal localization of the  
RT human pancreatic islet glucose-dependent insulinotropic polypeptide  
RT receptor.";  
RL Diabetes 44:1202-1208(1995).  
RN [4]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 96121393.  
RA YAMADA Y., HAYAMI T., NAKAMURA K., KAISAKI P.J., SOMEYA Y.,  
RA WANG C.Z., SEINO S., SEINO Y.;  
RT "Human gastric inhibitory polypeptide receptor: cloning of the gene  
RT (GIPR) and cDNA.";  
RL Genomics 29:773-776(1995).  
CC CC  
CC RECEPTOR IS MEDIATED BY G PROTEINS WHICH ACTIVATE ADENYL  
CYCLASE.  
CC CC  
CC SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.  
CC CC  
CC ALTERNATIVE PRODUCTS: TWO FORMS ARE PRODUCED BY ALTERNATIVE  
SPLICING.  
CC CC  
CC SIMILARITY: BELONGS TO FAMILY 2 OF G-PROTEIN COUPLED RECEPTORS.  
CC CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC CC  
DR EMBL; U39231; AAA84418.1;  
DR EMBL; S79852; AAB35419.1;  
DR EMBL; X81832; CAA57426.1;  
DR EMBL; D49559; BAA08503.1;  
DR EMBL; D49556; BAA08503.1; JOINED.  
DR EMBL; D49557; BAA08503.1; JOINED.

EMBL; D49558; BAA08503.1; JOINED.  
DR GCRDB; GCR\_1157;  
DR GCRDB; GCR\_1955;  
DR GCRDB; GCR\_1987;  
DR GCRDB; GCR\_2098;  
DR MIN; 137241;  
DR PROSITE; PS00649; G\_PROTEIN\_RECEP\_F2\_1; 1.  
DR PROSITE; PS00650; G\_PROTEIN\_RECEP\_F2\_2; 1.  
DR PFAM; PF00002; 7tm\_2; 1.  
KW G-protein coupled receptor; Transmembrane; Glycoprotein; Signal;  
KW Alternative splicing.  
FT SIGNAL 1 21 POTENTIAL.  
FT CHAIN 22 466 GASTRIC INHIBITORY POLYPEPTIDE RECEPTOR.  
FT DOMAIN 22 138 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 139 161 1 (POTENTIAL).  
FT DOMAIN 162 169 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 170 189 2 (POTENTIAL).  
FT DOMAIN 190 217 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 218 242 3 (POTENTIAL).  
FT DOMAIN 243 254 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 255 278 4 (POTENTIAL).  
FT DOMAIN 279 293 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 294 319 5 (POTENTIAL).  
FT DOMAIN 320 341 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 342 362 6 (POTENTIAL).  
FT DOMAIN 363 377 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 378 398 7 (POTENTIAL).  
FT DOMAIN 399 466 CYTOPLASMIC (POTENTIAL).  
FT CARBOHYD 62 62 POTENTIAL.  
FT CARBOHYD 77 77 POTENTIAL.  
FT VARSPLIC 399 399 V-> VGRDPAAPALWRRRGTTAPPLSAIVSQV (IN  
LONG ISOFORM).  
FT CONFLICT 12 12 R-> G (IN REF. 2).  
FT CONFLICT 104 104 G-> R (IN REF. 2).  
FT CONFLICT 117 117 MISSING (IN REF. 3).  
FT CONFLICT 337 337 L-> V (IN REF. 2).  
FT CONFLICT 367 371 GALRF->APCV (IN REF. 3).  
SQ SEQUENCE 466 AA; 53156 MW; 1DC57C17 CRC32;

Query Match 71.4%; Score 55; DB 1; Length 466;  
Best Local Similarity 77.8%; Pred. No. 2.30e+00;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 272 IPWIVRYL 280  
QY 1 IPFPIVRYL 9

RESULT 15  
ID CPl2\_MOUSE STANDARD; PRT; 513 AA.  
AC P00186;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 21-JUL-1986 (Rel. 01, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE CYTOCHROME P450 1A2 (EC 1.14.14.1) (CYP1A2) (P450-P2/P450-P3).  
GN CYP1A2 OR CYP1A-2.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
RN [1]  
RP SEQUENCE FROM N.A. (P3).  
RC STRAIN=C57BL/6N;  
RX MEDLINE: 84289486.  
RA KIMURA S., GONZALEZ F.J., NEBERT D.W.;  
RT "The murine Ah locus. Comparison of the complete cytochrome P1-450  
RT and P3-450 cDNA nucleotide and amino acid sequences.";  
RL J. Biol. Chem. 259:10705-10713(1984).  
RN [2]  
RP SEQUENCE FROM N.A. (P3).  
RC STRAIN=C57BL/6N;  
RX MEDLINE: 84169582.  
RA KIMURA S., GONZALEZ F.J., NEBERT D.W.;  
RT "Mouse cytochrome P3-450: complete cDNA and amino acid sequence.";

Db 131 FPIRYL 137  
||| |||  
QY 3 FPIRYL 9

Query Match 71.4%; Score 55; DB 1; Length 455;  
Best Local Similarity 77.8%; Pred. No. 2.30e+00;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

RESULT 12  
ID GIPR\_RAT STANDARD; PRT; 455 AA.  
AC P43219;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 35, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE GASTRIC INHIBITORY POLYPEPTIDE RECEPTOR PRECURSOR (GIP-R) (GLUCOSE-  
DEPENDENT INSULINOTROPIC POLYPEPTIDE RECEPTOR).  
GN GIPR.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-SPRAGUE-DAWLEY; TISSUE-BRAIN;  
RX MEDLINE; 94062667.  
RA USUIN T.B., MEZEY E., BUTTON D.C., BROWNSTEIN M.J., BONNER T.I.;  
RT "Gastric inhibitory polypeptide receptor, a member of the secretin-  
vasoactive intestinal peptide receptor family, is widely distributed  
in peripheral organs and the brain."  
RL Endocrinology 133:2861-2871(1993).  
CC -!- FUNCTION: THIS IS A RECEPTOR FOR GIP. THE ACTIVITY OF THIS  
RECEPTOR IS MEDIATED BY G PROTEINS WHICH ACTIVATE ADENYLIL  
CYCLASE.  
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.  
CC -!- TISSUE SPECIFICITY: PRESENT IN THE PANCREAS AS WELL AS THE GUT,  
ADIPOSE TISSUE, HEART, PITUITARY, AND INNER LAYERS OF THE ADRENAL  
CORTEX, WHEREAS IT IS NOT FOUND IN KIDNEY, SPLEEN, OR LIVER. IT IS  
ALSO EXPRESSED IN SEVERAL BRAIN REGIONS, INCLUDING THE CEREBRAL  
CORTEX, HIPPOCAMPUS, AND OLFACTORY BULB.  
CC -!- SIMILARITY: BELONGS TO FAMILY 2 OF G-PROTEIN COUPLED RECEPTORS.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; L19660; AAC37637.1; --  
DR GCRDB; GCR\_0817; --  
DR PROSITE; PS00649; G\_PROTEIN\_RECEPT\_F2\_1; 1.  
DR PROSITE; PS00650; G\_PROTEIN\_RECEPT\_F2\_2; 1.  
DR PFAM; PF00002; 7tm2; 1.  
KW G-protein coupled receptor; Transmembrane; Glycoprotein; Signal.  
FT SIGNAL 1 18  
FT CHAIN 19 455  
FT DOMAIN 19 135  
FT TRANSMEM 136 158  
FT DOMAIN 159 166  
FT TRANSMEM 167 186  
FT DOMAIN 187 214  
FT TRANSMEM 215 239  
FT DOMAIN 240 251  
FT TRANSMEM 252 275  
FT DOMAIN 276 290  
FT TRANSMEM 291 316  
FT DOMAIN 317 338  
FT TRANSMEM 339 359  
FT DOMAIN 360 374  
FT TRANSMEM 375 395  
FT DOMAIN 396 455  
FT CARBOHYD 59 59  
FT CARBOHYD 69 69  
FT CARBOHYD 74 74  
FT... SEQUENCE 455 AA; 52256 MW; A435EFB4 CRC32;

Db 269 IPWIVRYL 277  
||| |||  
QY 1 IPFPIVRYL 9

RESULT 13  
ID GIPR\_MESAU STANDARD; PRT; 462 AA.  
AC P43218;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE GASTRIC INHIBITORY POLYPEPTIDE RECEPTOR PRECURSOR (GIP-R) (GLUCOSE-  
DEPENDENT INSULINOTROPIC POLYPEPTIDE RECEPTOR).  
GN GIPR.  
OS Mesocricetus auratus (Golden hamster).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae; Mesocricetus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 95110292.  
RA YASUDA K., INAGAKI N., YAMADA Y., KUBOTA A., SEINO S., SEINO Y.;  
RT "Hamster gastric inhibitory polypeptide receptor expressed in  
pancreatic islets and clonal insulin-secreting cells: its structure  
and functional properties."  
RL Biochem. Biophys. Res. Commun. 205:1556-1562(1994).  
CC -!- FUNCTION: THIS IS A RECEPTOR FOR GIP. THE ACTIVITY OF THIS  
RECEPTOR IS MEDIATED BY G PROTEINS WHICH ACTIVATE ADENYLIL  
CYCLASE.  
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.  
CC -!- TISSUE SPECIFICITY: WIDELY DISTRIBUTED INCLUDING PANCREATIC  
ISLETS, BRAIN AND VARIOUS PERIPHERAL TISSUES.  
CC -!- SIMILARITY: BELONGS TO FAMILY 2 OF G-PROTEIN COUPLED RECEPTORS.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; D38103; BAA07284.1; --  
DR GCRDB; GCR\_1162; --  
DR PROSITE; PS00649; G\_PROTEIN\_RECEPT\_F2\_1; 1.  
DR PROSITE; PS00650; G\_PROTEIN\_RECEPT\_F2\_2; 1.  
DR PFAM; PF00002; 7tm2; 1.  
KW G-protein coupled receptor; Transmembrane; Glycoprotein; Signal.  
FT SIGNAL 1 18  
FT CHAIN 19 462  
FT DOMAIN 19 135  
FT TRANSMEM 136 158  
FT DOMAIN 159 166  
FT TRANSMEM 167 186  
FT DOMAIN 187 214  
FT TRANSMEM 215 239  
FT DOMAIN 240 251  
FT TRANSMEM 252 275  
FT DOMAIN 276 290  
FT TRANSMEM 291 316  
FT DOMAIN 317 338  
FT TRANSMEM 339 359  
FT DOMAIN 360 374  
FT TRANSMEM 375 395  
FT DOMAIN 396 462  
FT CARBOHYD 59 59  
FT CARBOHYD 74 74  
FT... SEQUENCE 462 AA; 52918 MW; 4A173782 CRC32;

QY 3 FPIVRYL 9  
:|||||

RESULT 9 PRT; 212 AA.  
ID YEAS-ECOLI STANDARD; PRT; 212 AA.  
AC P76249; 007971; 007969;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DE HYPOHETICAL 23.2 KD PROTEIN IN GAP-RND INTERGENIC REGION.  
GN YEAS.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Escherichia.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-K12 / MG1655;  
RX MEDLINE; 97426617.  
RA BLATTNER F.R., PLUNKETT G. III, BLOCH C.A., PERNA N.T., BURLAND V.,  
RA RILEY M., COLLADO-VIDES J., GLASNER F.D., RODE C.K., MAYHEW G.F.,  
RA GREGOR J., DAVIS N.W., KIRKPATRICK H.A., GOEDEN M.A., ROSE D.J.,  
RA MAU B., SHAO Y.;  
RT "The complete genome sequence of Escherichia coli K-12";  
RL Science 277:1453-1474(1997).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-K12;  
RX MEDLINE; 97251358.  
RA ITOH T., AIBA H., BABA T., FUJITA K., HAYASHI K., INADA T., ISONO K.,  
RA KASAI H., KIMURA S., KITAKAWA M., KITAGAWA M., MAKINO K., MIKI T.,  
RA MIZOBUCHI K., MORI H., MORI T., MOTOMURA K., NAKADE S., NAKAMURA Y.,  
RA NASHIMOTO H., NISHIO Y., OSHIMA T., SAITO N., SAMPEI G., SEKI Y.,  
RA SIVASUNDARAM S., TAGAMI H., TAKEDA J., TAKEMOTO K., WADA C.,  
RA YAMAMOTO Y., HORIUCHI T.;  
RT "A 460-kb DNA sequence of the Escherichia coli K-12 genome  
corresponding to the 40.1-50.0 min region on the linkage map";  
RL DNA Res. 3:379-392(1996).  
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).  
CC -1- SIMILARITY: BELONGS TO THE RHT FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

Query Match 71.4%; Score 55; DB 1; Length 212;  
Best Local Similarity 77.8%; Pred. No. 2.30e+00;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
DB 74 ILFNIVRYL 82  
QY 1 IPFPIVRYL 9  
:|||||

RESULT 10 PRT; 310 AA.  
ID KITH\_HSVTF STANDARD; PRT; 310 AA.  
AC P76249; 007971; 007969;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DE HYPOHETICAL 23.2 KD PROTEIN IN GAP-RND INTERGENIC REGION.  
GN YEAS.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Escherichia.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-K12 / MG1655;  
RX MEDLINE; 97426617.  
RA BLATTNER F.R., PLUNKETT G. III, BLOCH C.A., PERNA N.T., BURLAND V.,  
RA RILEY M., COLLADO-VIDES J., GLASNER F.D., RODE C.K., MAYHEW G.F.,  
RA GREGOR J., DAVIS N.W., KIRKPATRICK H.A., GOEDEN M.A., ROSE D.J.,  
RA MAU B., SHAO Y.;  
RT "The complete genome sequence of Escherichia coli K-12";  
RL Science 277:1453-1474(1997).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-K12;  
RX MEDLINE; 97251358.  
RA ITOH T., AIBA H., BABA T., FUJITA K., HAYASHI K., INADA T., ISONO K.,  
RA KASAI H., KIMURA S., KITAKAWA M., KITAGAWA M., MAKINO K., MIKI T.,  
RA MIZOBUCHI K., MORI H., MORI T., MOTOMURA K., NAKADE S., NAKAMURA Y.,  
RA NASHIMOTO H., NISHIO Y., OSHIMA T., SAITO N., SAMPEI G., SEKI Y.,  
RA SIVASUNDARAM S., TAGAMI H., TAKEDA J., TAKEMOTO K., WADA C.,  
RA YAMAMOTO Y., HORIUCHI T.;  
RT "A 460-kb DNA sequence of the Escherichia coli K-12 genome  
corresponding to the 40.1-50.0 min region on the linkage map";  
RL DNA Res. 3:379-392(1996).  
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).  
CC -1- SIMILARITY: BELONGS TO THE RHT FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

Query Match 71.4%; Score 55; DB 1; Length 212;  
Best Local Similarity 77.8%; Pred. No. 2.30e+00;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
DB 74 ILFNIVRYL 82  
QY 1 IPFPIVRYL 9  
:|||||

RESULT 10 PRT; 310 AA.  
ID KITH\_HSVTF STANDARD; PRT; 310 AA.  
AC P76249; 007971; 007969;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DE HYPOHETICAL 23.2 KD PROTEIN IN GAP-RND INTERGENIC REGION.  
GN YEAS.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Escherichia.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-K12 / MG1655;  
RX MEDLINE; 97426617.  
RA BLATTNER F.R., PLUNKETT G. III, BLOCH C.A., PERNA N.T., BURLAND V.,  
RA RILEY M., COLLADO-VIDES J., GLASNER F.D., RODE C.K., MAYHEW G.F.,  
RA GREGOR J., DAVIS N.W., KIRKPATRICK H.A., GOEDEN M.A., ROSE D.J.,  
RA MAU B., SHAO Y.;  
RT "The complete genome sequence of Escherichia coli K-12";  
RL Science 277:1453-1474(1997).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-K12;  
RX MEDLINE; 97251358.  
RA ITOH T., AIBA H., BABA T., FUJITA K., HAYASHI K., INADA T., ISONO K.,  
RA KASAI H., KIMURA S., KITAKAWA M., KITAGAWA M., MAKINO K., MIKI T.,  
RA MIZOBUCHI K., MORI H., MORI T., MOTOMURA K., NAKADE S., NAKAMURA Y.,  
RA NASHIMOTO H., NISHIO Y., OSHIMA T., SAITO N., SAMPEI G., SEKI Y.,  
RA SIVASUNDARAM S., TAGAMI H., TAKEDA J., TAKEMOTO K., WADA C.,  
RA YAMAMOTO Y., HORIUCHI T.;  
RT "A 460-kb DNA sequence of the Escherichia coli K-12 genome  
corresponding to the 40.1-50.0 min region on the linkage map";  
RL DNA Res. 3:379-392(1996).  
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).  
CC -1- SIMILARITY: BELONGS TO THE RHT FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

AC P13157;  
DT 01-JAN-1990 (Rel. 13, Created)  
DT 01-JAN-1990 (Rel. 13, Last sequence update)  
DT 01-NOV-1990 (Rel. 16, Last annotation update)  
DE THYMIDINE KINASE (EC 2.7.1.21).  
GN TK.  
OS Turkey herpesvirus (strain FC126).  
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 89259069.  
RA MARTIN S.L., APARISIO D.I., BANDYOPADHYAY P.K.;  
RT "Genetic and biochemical characterization of the thymidine kinase  
gene from herpesvirus of turkeys";  
RL J. Virol. 63:2847-2852(1989).  
CC -1- CATALYTIC ACTIVITY: ATP + THYMIDINE = ADP + THYMIDINE  
CC 5'-PHOSPHATE.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

EMBL; M26659; AAA46109.1; -  
DR PIR; A33346; KIBETH.  
DR HSSP; P03176; 1KIM.  
DR PFAM; PF00693; TK\_herpes; 1.  
KW Transferase; Kinase; DNA synthesis; ATP-binding.  
FT NP\_BIND 17 24 ATP (PROBABLE).  
SQ SEQUENCE 310 AA; 35512 MW; 927451D6 CRC32;

Query Match 71.4%; Score 55; DB 1; Length 310;  
Best Local Similarity 85.7%; Pred. No. 2.30e+00;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

DB 131 FPIVRYL 137  
QY 3 FPIVRYL 9  
:|||||

RESULT 11 STANDARD; PRT; 350 AA.  
ID KITH\_HSVTU  
AC P25987;  
DT 01-MAY-1992 (Rel. 22, Created)  
DT 01-MAY-1992 (Rel. 22, Last sequence update)  
DE THYMIDINE KINASE (EC 2.7.1.21).  
GN TK.  
OS Turkey herpesvirus.  
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 90063552.  
RA SCOTT S.D., ROSS N.L.J., BINNS M.M.;  
RT "Nucleotide and predicted amino acid sequences of the Marek's disease  
virus and turkey herpesvirus thymidine kinase genes; comparison with  
thymidine kinase genes of other herpesviruses";  
RL J. Gen. Virol. 70:3055-3065(1989).  
CC -1- CATALYTIC ACTIVITY: ATP + THYMIDINE = ADP + THYMIDINE  
CC 5'-PHOSPHATE.  
CC PIR; A33375; KIBEFC.  
DR HSSP; P03176; 1KIM.  
DR PFAM; PF00693; TK\_herpes; 1.  
KW Transferase; Kinase; DNA synthesis; ATP-binding.  
FT NP\_BIND 17 24 ATP (PROBABLE).  
SQ SEQUENCE 350 AA; 39968 MW; CB4E471B CRC32;

Query Match 71.4%; Score 55; DB 1; Length 350;  
Best Local Similarity 85.7%; Pred. No. 2.30e+00;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

DB 131 FPIVRYL 137  
QY 3 FPIVRYL 9  
:|||||

RESULT 11 STANDARD; PRT; 350 AA.  
ID KITH\_HSVTU  
AC P25987;  
DT 01-MAY-1992 (Rel. 22, Created)  
DT 01-MAY-1992 (Rel. 22, Last sequence update)  
DE THYMIDINE KINASE (EC 2.7.1.21).  
GN TK.  
OS Turkey herpesvirus.  
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 90063552.  
RA SCOTT S.D., ROSS N.L.J., BINNS M.M.;  
RT "Nucleotide and predicted amino acid sequences of the Marek's disease  
virus and turkey herpesvirus thymidine kinase genes; comparison with  
thymidine kinase genes of other herpesviruses";  
RL J. Gen. Virol. 70:3055-3065(1989).  
CC -1- CATALYTIC ACTIVITY: ATP + THYMIDINE = ADP + THYMIDINE  
CC 5'-PHOSPHATE.  
CC PIR; A33375; KIBEFC.  
DR HSSP; P03176; 1KIM.  
DR PFAM; PF00693; TK\_herpes; 1.  
KW Transferase; Kinase; DNA synthesis; ATP-binding.  
FT NP\_BIND 17 24 ATP (PROBABLE).  
SQ SEQUENCE 350 AA; 39968 MW; CB4E471B CRC32;

Query Match 71.4%; Score 55; DB 1; Length 350;  
Best Local Similarity 85.7%; Pred. No. 2.30e+00;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

DB 131 FPIVRYL 137  
QY 3 FPIVRYL 9  
:|||||



-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----

DR EMBL; Z00036; CAA77335.1; -;  
DR EMBL; M31667; AAA52163.1; -;  
DR EMBL; M31664; AAA52163.1; JOINED.  
DR EMBL; M31665; AAA52163.1; JOINED.  
DR EMBL; M31666; AAA52163.1; JOINED.  
DR EMBL; M12078; AAA52154.1; -;  
DR EMBL; L00389; AAA35738.1; -;  
DR EMBL; L00384; AAA35738.1; JOINED.  
DR EMBL; L00385; AAA35738.1; JOINED.  
DR EMBL; L00386; AAA35738.1; JOINED.  
DR EMBL; L00388; AAA35738.1; JOINED.  
DR EMBL; L00387; AAA35738.1; JOINED.  
DR EMBL; M55053; AAA52146.1; -;  
DR PIR; S07373; O4HU4.  
DR PIR; S16718; S16718.  
DR PIR; S22433; S22433.  
DR MIM; 124060; -;  
DR MIM; 108330; -;  
DR PROSITE; PS00086; CYTOCHROME\_P450; 1.  
DR PFAM; PF00067; p450; 1.  
KW Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme;  
KW Microsome; Endoplasmic reticulum.  
FT BINDING 458 458 HEME (BY SIMILARITY).  
FT CONFLICT 79 79 R -> S (IN REF. 4).  
FT CONFLICT 311 311 V -> L (IN REF. 5).  
FT CONFLICT 450 451 LF -> MLV (IN REF. 5).  
FT CONFLICT 511 511 R -> LP (IN REF. 4).  
SQ SEQUENCE 515 AA; 58294 MW; C87C0B4C CRC32;

Query Match 72.7%; Score 56; DB 1; Length 515;  
Best Local Similarity 85.7%; Pred. No. 1.44e+00;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
Db 239 PFILRYL 245  
Qy 3 PFIVRYL 9  
|||||

RESULT 5  
ID CP12\_CAVPO STANDARD; PRT; 515 AA.  
AC Q64391; Q64404;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 01-JUL-1999 (Rel. 38, Last annotation update)  
DE CYTOCHROME P450 1A2 (EC 1.14.14.1) (CYP1A2).  
GN CYP1A2.  
OS Cavia porcellus (Guinea pig).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Rodentia; Hystricognathi; Caviidae; Cavia.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-HARTLEY; TISSUE=LIVER;  
RX MEDLINE; 98096351.  
RA MORI T., ITOH S., OHGIYA S., ISHIZAKI K., KAMATAKI T.;  
RT "Regulation of CYP1A and CYP3A mRNAs by ascorbic acid in guinea  
RT pigs."  
RL Arch. Biochem. Biophys. 348:268-277 (1997).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-HARTLEY; TISSUE=LIVER;  
RA BLACK V.H.;  
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: CYTOCHROMES P450 ARE A GROUP OF HEME-THIOLATE  
CC MONOOXYGENASES. IN LIVER MICROSOMES, THIS ENZYME IS INVOLVED IN AN.

CC NADPH-DEPENDENT ELECTRON TRANSPORT PATHWAY. IT OXIDIZES A VARIETY  
CC OF STRUCTURALLY UNRELATED COMPOUNDS, INCLUDING STEROIDS, FATTY  
CC ACIDS, AND XENOBIOTICS.  
CC -!- CATALYTIC ACTIVITY: RH + REDUCED FLAVOPROTEIN + O(2) -> ROH +  
CC OXIDIZED FLAVOPROTEIN + H(2)O.  
CC -!- SUBCELLULAR LOCATION: MEMBRANE-BOUND. ENDOPLASMIC RETICULUM.  
CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----

DR EMBL; D50457; BAA09048.1; -;  
DR EMBL; U23501; AAB70866.1; -;  
DR PROSITE; PS00086; CYTOCHROME\_P450; 1.  
DR PFAM; PF00067; p450; 1.  
KW Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme;  
KW Microsome; Endoplasmic reticulum.  
FT BINDING 458 458 HEME (BY SIMILARITY).  
FT CONFLICT 149 149 V -> L (IN REF. 2).  
SQ SEQUENCE 515 AA; 58422 MW; 93E6811E CRC32;

Query Match 72.7%; Score 56; DB 1; Length 515;  
Best Local Similarity 85.7%; Pred. No. 1.44e+00;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
Db 239 PFILRYL 245  
Qy 3 PFIVRYL 9  
|||||

RESULT 6  
ID CP11\_CANFA STANDARD; PRT; 524 AA.  
AC P56590;  
DT 15-DEC-1998 (Rel. 37, Created)  
DT 15-DEC-1998 (Rel. 37, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE CYTOCHROME P450 1A1 (EC 1.14.14.1) (CYP1A1) (DAH1).  
GN CYP1A1.  
OS Canis familiaris (Dog).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Carnivora; Fissipedia; Canidae; Canis.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BEAGLE; TISSUE=LIVER;  
RX MEDLINE; 91042464.  
RA UCHIDA T., KOMORI M., KITADA M., KAMATAKI T.;  
RT "Isolation of cDNAs coding for three different forms of liver  
RT microsomal cytochrome P-450 from polychlorinated biphenyl-treated  
RT beagle dogs."  
RL Mol. Pharmacol. 38:644-651 (1990).  
CC -!- FUNCTION: CYTOCHROMES P450 ARE A GROUP OF HEME-THIOLATE  
CC MONOOXYGENASES. IN LIVER MICROSOMES, THIS ENZYME IS INVOLVED IN AN  
CC NADPH-DEPENDENT ELECTRON TRANSPORT PATHWAY. THIS ISOZYME SEEMS  
CC RESPONSIBLE FOR METABOLISM OF 2',4',4',5',5'-HEXACHLOROBIPHENYL.  
CC -!- CATALYTIC ACTIVITY: RH + REDUCED FLAVOPROTEIN + O(2) -> ROH +  
CC OXIDIZED FLAVOPROTEIN + H(2)O.  
CC -!- SUBCELLULAR LOCATION: MEMBRANE-BOUND. ENDOPLASMIC RETICULUM.  
CC -!- INDUCTION: BY POLYCHLORINATED BIPHENYL (PCB) IN LIVER AND KIDNEY.  
CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.  
CC -!- CAUTION: IT IS UNCERTAIN WHETHER MET-1, MET-2 OR MET-4 IS THE  
CC INITIATOR.  
DR PROSITE; PS00086; CYTOCHROME\_P450; 1.  
DR PFAM; PF00067; p450; 1.  
KW Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme;  
KW Microsome; Endoplasmic reticulum.  
FT BINDING 461 461 HEME (BY SIMILARITY).  
SQ SEQUENCE 524 AA; 59209 MW; CECAC33B CRC32;

RA JOHNSON E.F., TUKEY R.H.;  
RT "Cloning and characterization of cDNAs encoding 2,3,7,8-  
RT tetrachlorodibenzo-p-dioxin-inducible rabbit mRNAs for cytochrome  
RT P-450 isozymes 4 and 6."  
RL Proc. Natl. Acad. Sci. U.S.A. 82:5310-5314(1985).  
RN [5]  
RN PARTIAL SEQUENCE.  
RX MEDLINE; 84272618.  
RA FUJITA V.S., BLACK S.D., TARR G.E., KOOP D.R., COON M.J.;  
RT "On the amino acid sequence of cytochrome P-450 isozyme 4 from rabbit  
RT liver microsomes."  
RL Proc. Natl. Acad. Sci. U.S.A. 81:4260-4264(1984).  
CC -1- FUNCTION: CYTOCHROMES P450 ARE A GROUP OF HEME-THIOLATE  
CC MONOOXYGENASES. IN LIVER MICROSOMES, THIS ENZYME IS INVOLVED IN AN  
CC NADPH-DEPENDENT ELECTRON TRANSPORT PATHWAY. IT OXIDIZES A VARIETY  
CC OF STRUCTURALLY UNRELATED COMPOUNDS, INCLUDING STEROIDS, FATY  
CC ACIDS, AND XENOBIOTICS.  
CC -1- CATALYTIC ACTIVITY: RH + REDUCED FLAVOPROTEIN + O(2) = ROH +  
CC OXIDIZED FLAVOPROTEIN + H(2)O.  
CC -1- SUBCELLULAR LOCATION: MEMBRANE-BOUND. ENDOPLASMIC RETICULUM.  
CC -1- INDUCTION: BY BETA-NAPHTHOFLAVONE AND BY 2,3,7,8-  
CC TETRACHLORODIBENZO-P-DIOXIN (TCDD).  
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.  
CC  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; X13853; CAA32066.1; -  
CC EMBL; M36538; AAA31437.1; -  
CC EMBL; M1728; AAA31433.1; ALT\_SEQ.  
CC EMBL; D00213; BAA00152.1; -  
CC EMBL; X05686; CAA29171.1; -  
CC PIR; A00187; O4RBBN.  
CC PIR; A00188; A00188.  
CC PIR; S02038; S02038.  
CC PIR; B25143; B25143.  
CC PIR; B27821; B27821.  
CC PFAM; PF00067; CYTOCHROME\_P450; 1.  
CC PROSITE; PS00086; CYTOCHROME\_P450; 1.  
CC Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme;  
KW Microsome; Endoplasmic reticulum.  
FT INIT\_MET 0  
FT BINDING 360 364  
FT ACT\_SITE 402 402  
FT  
FT BINDING 457 457  
FT VARIANT 173 173  
FT VARIANT 232 232  
FT VARIANT 298 298  
FT CONFLICT 21 21  
FT CONFLICT 66 66  
FT CONFLICT 69 69  
FT CONFLICT 91 91  
FT CONFLICT 120 120  
FT CONFLICT 171 171  
FT CONFLICT 193 193  
FT CONFLICT 207 207  
FT CONFLICT 246 246  
FT CONFLICT 250 250  
FT CONFLICT 288 301  
FT  
FT CONFLICT 353 354  
FT CONFLICT 357 357  
FT CONFLICT 358 358  
FT CONFLICT 461 461  
FT CONFLICT 493 493  
FT SEQUENCE 515 AA; 58202 MW; 552A615C CRC32;

Query Match 72.7%; Score 56; DB 1; Length 515;  
Best Local Similarity 85.7%; Pred. No. 1.44e+00;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
Db 238 FPIRLYL 244  
QY 3 FPIRVYL 9  
RESULT 4  
ID CP12\_HUMAN STANDARD; PRT; 515 AA.  
AC P05177; Q16754;  
DT 13-AUG-1987 (Rel. 05, Created)  
DT 13-AUG-1987 (Rel. 05, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE CYTOCHROME P450 1A2 (EC 1.14.14.1) (CYP1A2) (P450-P3) (P(3)450) (P450  
DE 4).  
GN Homo sapiens (Human).  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 86312938.  
RA JAISWAL A.K., NEBERT D.W., GONZALEZ F.J.;  
RT "Human P3(450): cDNA and complete amino acid sequence."  
RL Nucleic Acids Res. 14:6773-6774(1986).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 90114205.  
RA IKEYA K., JAISWAL A.K., OWENS R.A., JONES J.E., NEBERT D.W.,  
RA KIMURA S.;  
RT "Human CYP1A2: sequence, gene structure, comparison with the mouse  
RT and rat orthologous gene, and differences in liver 1A2 mRNA  
RT expression."  
RL Mol. Endocrinol. 3:1399-1408(1989).  
RN [3]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 88061719.  
RA JAISWAL A.K., NEBERT D.W., MCBRIDE O.W., GONZALEZ F.J.;  
RT "Human P(3)450: cDNA and complete protein sequence, repetitive Alu  
RT sequences in the 3' nontranslated region, and localization of gene to  
RT chromosome 15."  
RL J. Exp. Pathol. 3:1-17(1987).  
RN [4]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 86313652.  
RA QUATTROCHI L.C., PENDURTHI U.R., OKINO S.T., POTENZA C., TUKEY R.H.;  
RT "Human cytochrome P-450 4 mRNA and gene: part of a multigene family  
RT that contains Alu sequences in its mRNA."  
RL Proc. Natl. Acad. Sci. U.S.A. 83:6731-6735(1986).  
RN [5]  
RP SEQUENCE OF 295-485 FROM N.A.  
RX MEDLINE; 86081170.  
RA QUATTROCHI L.C., OKINO S.T., PENDURTHI U.R., TUKEY R.H.;  
RT "Cloning and isolation of human cytochrome P-450 cDNAs homologous to  
RT dioxin-inducible rabbit mRNAs encoding P-450 4 and P-450 6."  
RL DNA 4:395-400(1985).  
CC -1- FUNCTION: CYTOCHROMES P450 ARE A GROUP OF HEME-THIOLATE  
CC MONOOXYGENASES. IN LIVER MICROSOMES, THIS ENZYME IS INVOLVED IN AN  
CC NADPH-DEPENDENT ELECTRON TRANSPORT PATHWAY. IT OXIDIZES A VARIETY  
CC OF STRUCTURALLY UNRELATED COMPOUNDS, INCLUDING STEROIDS, FATY  
CC ACIDS, AND XENOBIOTICS.  
CC -1- CATALYTIC ACTIVITY: RH + REDUCED FLAVOPROTEIN + O(2) = ROH +  
CC OXIDIZED FLAVOPROTEIN + H(2)O.  
CC -1- SUBCELLULAR LOCATION: MEMBRANE-BOUND. ENDOPLASMIC RETICULUM.  
CC -1- TISSUE SPECIFICITY: LIVER.  
CC -1- INDUCTION: P450 CAN BE INDUCED TO HIGH LEVELS IN LIVER AND OTHER  
CC TISSUES BY VARIOUS FOREIGN COMPOUNDS, INCLUDING DRUGS, PESTICIDES,  
CC AND CARCINOGENS.  
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.



CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: M15349; NOT\_ANNOTATED\_CDS.  
DR EMBL: D84432; BAA12661.1; -  
DR EMBL: L09228; AAA67472.1; -  
DR EMBL: D90189; BAA14210.1; -  
DR EMBL: Z99116; CAB14271.1; -  
DR PIR: PS0430; PS0430.  
DR SUBTILIST: BG10508; SPOVAF.  
KW Sporulation; Transmembrane.  
FT TRANSMEM 106 126 POTENTIAL.  
FT TRANSMEM 252 272 POTENTIAL.  
FT TRANSMEM 296 316 POTENTIAL.  
FT TRANSMEM 335 355 POTENTIAL.  
FT TRANSMEM 363 383 POTENTIAL.  
FT TRANSMEM 387 407 POTENTIAL.  
FT CONFLICT 432 432 L -> I (IN REF. 2).  
FT CONFLICT 467 467 R -> P (IN REF. 4).  
SQ SEQUENCE 492 AA; 55606 MW; 6CBA2C32 CRC32;

Query Match 74.0%; Score 57; DB 1; Length 492;  
Best Local Similarity 85.7%; Pred. No. 8.97e-01;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 232 PFPLVRY 238  
|||:||||  
QY 2 PFPIVRY 8

RESULT 2  
ID CP12\_MESAU STANDARD; PRT; 513 AA.  
AC P24453;  
DT 01-MAR-1992 (Rel. 21, Created)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE CYTOCHROME P450 1A2 (EC 1.14.14.1) (CYP1A2) (P450-MC4) (HEPATIC  
DE CYTOCHROME P-450MC1).  
GN CYP1A2.  
OS Mesocricetus auratus (Golden hamster).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae; Mesocricetus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=LIVER;  
RX MEDLINE; 91112759.  
RA LAI T.S., CHIANG J.Y.L.;  
RT "Cloning and characterization of two major 3-methylcholanthrene  
RT inducible hamster liver cytochrome P450s.";  
RL Arch. Biochem. Biophys. 283:429-439(1990).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 92138673.  
RA SAGAMI I., OHMACHI T., FUJII H., KIKUCHI H., WATANABE M.;  
RT "Hamster cytochrome P-450 1A gene family, P-4501A1 and P-4501A2 in  
RT lung and liver: cDNA cloning and sequence analysis.";  
RL J. Biochem. 110:641-647(1991).  
CC -1- FUNCTION: CYTOCHROMES P450 ARE A GROUP OF HEME-THIOLATE  
CC MONOOXYGENASES. IN LIVER MICROSOMES, THIS ENZYME IS INVOLVED IN AN  
CC NADPH-DEPENDENT ELECTRON TRANSPORT PATHWAY. IT OXIDIZES A VARIETY  
CC OF STRUCTURALLY UNRELATED COMPOUNDS, INCLUDING STEROIDS, FATTY  
CC ACIDS, AND XENOBIOTICS.  
CC -1- CATALYTIC ACTIVITY: RH + REDUCED FLAVOPROTEIN + O(2) -> ROH +  
CC OXIDIZED FLAVOPROTEIN + H(2)O.  
CC -1- SUBCELLULAR LOCATION: MEMBRANE-BOUND. ENDOPLASMIC RETICULUM.  
CC -1- TISSUE SPECIFICITY: FOUND IN LUNG AND LIVER.  
CC -1- INDUCTION: BY 3-METHYLCHOLANTHRENE (3MC).  
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.  
CC  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: M63787; AAA37070.1; -  
DR EMBL: D10252; BAA01097.1; -  
DR EMBL: D10914; BAA01718.1; -  
DR PIR: S13885; S13885.  
DR PROSITE: PS00086; CYTOCHROME\_P450; 1.  
DR PFAM: PF00067; p450; 1.  
KW Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme;  
KW Microsome; Endoplasmic reticulum.  
FT BINDING 456 456 HEME (BY SIMILARITY).  
FT CONFLICT 49 49 I -> F (IN REF. 1).  
FT CONFLICT 52 53 HV -> MC (IN REF. 1).  
FT CONFLICT 253 254 KN -> GG (IN REF. 1).  
FT CONFLICT 326 326 L -> W (IN REF. 1).  
FT CONFLICT 356 356 R -> L (IN REF. 1).  
FT CONFLICT 485 485 T -> Q (IN REF. 1).  
SQ SEQUENCE 513 AA; 58082 MW; 40F0041D CRC32;

Query Match 72.7%; Score 56; DB 1; Length 513;  
Best Local Similarity 85.7%; Pred. No. 1.44e+00;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 238 FPILRYL 244  
|||:||||  
QY 3 FPIVRYL 9

RESULT 3  
ID CP12\_RABIT STANDARD; PRT; 515 AA.  
AC P00187; Q29526;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 01-JAN-1990 (Rel. 13, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE CYTOCHROME P450 1A2 (EC 1.14.14.1) (CYP1A2) (P450 ISOZYME 4) (P450-  
DE PM4) (P450 LM4).  
GN CYP1A2.  
OS Oryctolagus cuniculus (Rabbit).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Lagomorpha; Leporidae; Oryctolagus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NEW ZEALAND WHITE;  
RX MEDLINE; 89052697.  
RA POMPON D.;  
RT "cDNA cloning and functional expression in yeast Saccharomyces  
RT cerevisiae of beta-naphthoflavone-induced rabbit liver P-450 LM4 and  
RT LM6.";  
RL Eur. J. Biochem. 177:285-293(1988).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=LIVER;  
RX MEDLINE; 88032911.  
RA KAGAWA N., MIHARA K., SATO R.;  
RT "Structural analysis of cloned cDNAs for polycyclic hydrocarbon-  
RT inducible forms of rabbit liver microsomal cytochrome P-450.";  
RL J. Biochem. 101:1471-1479(1987).  
RN [3]  
RP SEQUENCE.  
RX MEDLINE; 86140205.  
RA OZOLS J.;  
RT "Complete amino acid sequence of a cytochrome P-450 isolated from  
RT beta-naphthoflavone-induced rabbit liver microsomes. Comparison with  
RT phenobarbital-induced and constitutive isozymes and identification of  
RT invariant residues.";  
RL J. Biol. Chem. 261:3965-3979(1986).  
RN [4]  
RP SEQUENCE OF 91-514 FROM N.A.  
RX MEDLINE; 85270514.  
RA OKINO S.T., QUATTROCHI L.C., BARNES H.J., OSANTO S., GRIFFIN K.J.,

[3] SEQUENCE OF 81-492 FROM N.A.  
 RC STRAIN-168 / MARBURG;  
 MEDLINE; 95020538.  
 X SOROKIN A.V., ZUMSTEIN E., AZEVEDO V., EHRLICH S.D., SERROR P.;  
 X "The organization of the Bacillus subtilis 168 chromosome region  
 X between the spoVA and serA genetic loci, based on sequence data.";  
 X Mol. Microbiol. 10:385-395(1993).  
 X [4]  
 X SEQUENCE OF 223-492 FROM N.A.  
 X MEDLINE; 91345841.  
 X YAMAMOTO J., SHIMIZU M., YAMANE K.;  
 X "Molecular cloning and analysis of nucleotide sequence of the  
 X Bacillus subtilis lysa gene region using B. subtilis phage vectors  
 X and a multi-copy plasmid, pUBI10.";  
 X Agric. Biol. Chem. 55:1615-1626(1991).  
 X -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (PROBABLE).  
 X -1- SIMILARITY: SOME, TO SPORE GERMINATION PROTEIN GERAA AND GERBA.  
 X -----  
 X THIS SWISS-PROT entry is copyright. It is produced through a collabora-  
 X between the Swiss Institute of Bioinformatics and the EMBL outstation  
 X the European Bioinformatics Institute. There are no restrictions on  
 X use by non-profit institutions as long as its content is in no  
 X way to be reproduced for commercial purposes without the written  
 X permission of the EMBL outstation.

THIS PAGE BLANK (USTPO)

RP SEQUENCE FROM N.A.  
RC STRAIN-VF5;  
RA DECKERT G., WARREN P.V., GAASTERLAND T., YOUNG W.G., LENOX A.L.,  
RA GRAHAM D.E., OVERBEK R., SNEAD M.A., KELLER M., AUJAY M., HUBER R.,  
RA FELDMAN R.A., SHORT J.M., OLSON G.J., SWANSON R.V.;  
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AE006672; AAC06449.1; -;  
KW Hypothetical protein.  
SQ SEQUENCE 200 AA; 23046 MW; 20F220E5 CRC32;  
  
Query Match 68.8%; Score 53; DB 2; Length 200;  
Best Local Similarity 55.6%; Pred. No. 1.55e+01;  
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
  
Db 101 VPFPIKPYL 109  
:||||:|  
QY 1 IPFFIVRYL 9  
  
RESULT 13  
ID P92010 PRELIMINARY; PRT; 588 AA.  
AC P92010;  
DT 01-MAY-1997 (TrEMBLrel. 03, Created)  
DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)  
DT 01-JAN-1999 (TrEMBLrel. 09, Last annotation update)  
DE R10D12.8 PROTEIN.  
GN R10D12.8.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;  
OC Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA PERCY C.;  
RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 94150718.  
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,  
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,  
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,  
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIEZ M., JOHNSTON L.,  
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,  
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,  
RA PARSONS J., PERCY C., RIEKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,  
RA SWALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,  
RA THERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,  
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WORLDMAN P.;  
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
RT elegans.";  
RL Nature 368:32-38(1994).  
DR EMBL; 281109; CAB03251.1; -;  
SQ SEQUENCE 588 AA; 69273 MW; 697E93DF CRC32;  
  
Query Match 68.8%; Score 53; DB 5; Length 588;  
Best Local Similarity 55.6%; Pred. No. 1.55e+01;  
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;  
  
Db 108 VSEFINREL 116  
:||||:|  
QY 1 IPFFIVRYL 9

RESULT 14  
ID O54767 PRELIMINARY; PRT; 695 AA.  
AC O54767;  
DT 01-JUN-1998 (TrEMBLrel. 06, Created)  
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)  
DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)  
DE ZONA PELLUCIDA 2 GLYCOPROTEIN.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
RN [1]

RP SEQUENCE FROM N.A.  
RC STRAIN-SPRAGUE-DAWLEY; TISSUE-OVARY;  
RA AKATSUKA K., ARAKI Y.;  
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB000929; BAA24487.1; -;  
DR PFAM; PF00100; zona\_pellucida; 1.  
DR PRINTS; PR00023; ZPELUCIDA.  
SQ SEQUENCE 695 AA; 78443 MW; EF633556 CRC32;  
  
Query Match 68.8%; Score 53; DB 11; Length 695;  
Best Local Similarity 71.4%; Pred. No. 1.55e+01;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
  
Db 496 YPLVRYL 502  
:||||  
QY 3 FPIVRYL 9  
  
RESULT 15  
ID O76151 PRELIMINARY; PRT; 948 AA.  
AC O76151;  
DT 01-NOV-1998 (TrEMBLrel. 08, Created)  
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)  
DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)  
DE AMINOPEPTIDASE N.  
GN APN2.  
OS Bombyx mori (Silk moth).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;  
OC Bombycoidea; Bombycidae; Bombyx.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-KINSHU X SHOWA; TISSUE-MIDGUT;  
RA GANG H., TSUKAMOTO K., IKEZAWA H.;  
RT "Cloning and sequence analysis of the aminopeptidase N isoform(apn2)  
RT cDNA from Bombyx mori midgut.";  
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB011497; BAA32140.1; -;  
DR PFAM; PF01433; Peptidase\_M1; 1.  
DR PRINTS; PR00756; ALADIPTASE.  
KW Aminopeptidase.  
SQ SEQUENCE 948 AA; 107480 MW; 75BA2054 CRC32;  
  
Query Match 68.8%; Score 53; DB 5; Length 948;  
Best Local Similarity 66.7%; Pred. No. 1.55e+01;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
  
Db 632 LPFEISRYL 640  
:||||  
QY 1 IPFFIVRYL 9

Search completed: Fri Apr 14 23:58:37 2000  
Job time : 102 secs.

KW Transferase; Kinase; DNA synthesis; ATP-binding.  
FT NP\_BIND 19 26 ATP (POTENTIAL).  
SQ SEQUENCE 341 AA; 37956 MW; CE0C9B1C CRC32;

Query Match 70.1%; Score 54; DB 14; Length 341;  
Best Local Similarity 66.7%; Pred. No. 1.00e-01;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 137 IRFPLSKYL 145

QY 1 IPFIVRYL 9

RESULT 9  
ID Q09998 PRELIMINARY; PRT; 492 AA.

AC Q09998;  
DT 01-NOV-1996 (TREMELrel. 01, Created)  
DT 01-NOV-1996 (TREMELrel. 01, Last sequence update)  
DT 01-JAN-1998 (TREMELrel. 05, Last annotation update)  
DE PUTATIVE 55.5 KD ZINC FINGER PROTEIN R144.3 IN CHROMOSOME III.  
GN R144.3.

OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;  
OC Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.  
RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-BRISTOL N2;

RA FAVELLO T.;

RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.

CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).

DR EMBL; U23515; AAC46547.1; -.

DR WORMPEP; R144.3; CE02033.

DR PROSITE; PS00028; ZINC\_FINGER\_C2H2; 2.

KW Hypothetical protein; Zinc-finger; DNA-binding; Metal-binding;

FT DOMAIN 230 287 ZINC-FINGERS.

FT ZN\_FING 228 233 C2H2-TYPE.

FT ZN\_FING 264 287 C2H2-TYPE.

SQ SEQUENCE 492 AA; 55479 MW; D62CA443 CRC32;

Query Match 70.1%; Score 54; DB 5; Length 492;

Best Local Similarity 85.7%; Pred. No. 1.00e-01;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 109 FOIVRYL 115

QY 3 FPIVRYL 9

RESULT 10  
ID Q17457 PRELIMINARY; PRT; 189 AA.

AC Q17457;

DT 01-NOV-1996 (TREMELrel. 01, Created)

DT 01-NOV-1996 (TREMELrel. 01, Last sequence update)

DT 01-JAN-1999 (TREMELrel. 09, Last annotation update)

DE B0284.3 PROTEIN.

GN B0284.3.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;

OC Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.

RN [1]

RP SEQUENCE FROM N.A.

RA SULSTON J.;

RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE; 94150718.

RA WILSON R.; AINSCOUGH R.; ANDERSON K.; BAYNES C.; BERKS M.;

RA BONFIELD J.; BURTON J.; CONNELL M.; COPSEY T.; COOPER J.; COULSON A.;

RA CRAXTON M.; DEAR S.; DU Z.; DURBIN R.; FAVELLO A.; FULTON L.;

RA GARDNER A.; GREEN P.; HAWKINS T.; HILLIER L.; JIER M.; JOHNSTON L.;

RA JONES M.; KERSHAW J.; KIRSTEN J.; LAISTER N.; LATREILLE P.;

RA LIGHTNING J.; LLOYD C.; MCMURRAY A.; MORTIMORE B.; O'CALLAGHAN M.;

RA PARSONS J.; PERCY C.; RIFKEN L.; ROOPRA A.; SAUNDERS D.; SHOWNKEEN R.;

RA SMALDON N.; SMITH A.; SONNHAMMER E.; STADEN R.; SULSTON J.;

RA THIERRY-MIEG J.; THOMAS K.; VAUDIN M.; VAUGHAN K.; WATERSTON R.;

RA WATSON A.; WEINSTOCK L.; WILKINSON-SPROAT J.; WOHLDMAN P.;

RT "2.2 kb of contiguous nucleotide sequence from chromosome III of C. elegans.";

RL Nature 368:32-38(1994).

DR EMBL; Z30973; CAA83221.1; -.

SQ SEQUENCE 189 AA; 21347 MW; C31E7906 CRC32;

Query Match 68.8%; Score 53; DB 5; Length 189;  
Best Local Similarity 75.0%; Pred. No. 1.55e-01;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 23 PFIMRYL 30

QY 2 PPIVRYL 9

RESULT 11

ID Q94366 PRELIMINARY; PRT; 192 AA.

AC Q94366;

DT 01-MAY-1999 (TREMELrel. 10, Created)

DT 01-MAY-1999 (TREMELrel. 10, Last sequence update)

DT 01-NOV-1999 (TREMELrel. 12, Last annotation update)

DE ANTHRANILATE SYNTHASE SMALL SUBUNIT.

GN TRPG.

OS Buchnera aphidicola.

OG Plasmid pBps2.

OC Bacteria; Proteobacteria; gamma subdivision; Buchnera.

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE; 99091362.

RA VAN HAM R.C.; MARTINEZ-TORRES D.; MOYA A.; LATORRE A.;

RT "Plasmid-encoded anthranilate synthase (TrpEG) in Buchnera aphidicola

from aphids of the family Pemphigidae.";

RL Appl. Environ. Microbiol. 65:117-125(1999).

DR EMBL; AJ012334; CAA10000.1; -.

DR EMBL; AJ012334; CAA09997.1; -.

DR PROSITE; PS00442; GATASE\_TYPE\_I; 1.

KW Plasmid.

SQ SEQUENCE 192 AA; 21668 MW; E0ACC5AD CRC32;

Query Match 68.8%; Score 53; DB 2; Length 192;

Best Local Similarity 71.4%; Pred. No. 1.55e-01;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 126 PFPVARY 132

QY 2 PPIVRY 8

RESULT 12

ID O66494 PRELIMINARY; PRT; 200 AA.

AC O66494;

DT 01-AUG-1998 (TREMELrel. 07, Created)

DT 01-AUG-1998 (TREMELrel. 07, Last sequence update)

DT 01-NOV-1998 (TREMELrel. 08, Last annotation update)

DE HYPOTHETICAL 23.0 KD PROTEIN.

GN AQ.082.

OS Aquifex aeolicus.

OC Bacteria; Aquificales; Aquificaceae; Aquifex.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-VF5;

RX MEDLINE; 98196666.

RA DECKERT G.; WARREN P.V.; GAASTERLAND T.; YOUNG W.G.; LENOX A.L.;

RA GRAHAM D.E.; OVERBECK R.; SNEAD M.A.; KELLER M.; AUJAY M.; HUBER R.;

RA FELDMAN R.A.; SHORT J.M.; OLSON G.J.; SWANSON R.V.;

RT "The complete genome of the hyperthermophilic bacterium Aquifex

aeolicus.";

RL Nature 392:353-358(1998).

RN [2]

Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.

```

RL  Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
RN  [2]
RP  SEQUENCE FROM N.A.
RC  STRAIN-JH642;
RX  MEDLINE; 96074318.
RA  LIN Y., HANSEN J.N.;
RT  "Characterization of a chimeric prov operon in a subtilin-producing
RT  mutant of Bacillus subtilis 168."
RL  J. Bacteriol. 177:6874-6880(1995).
RN  [3]
RP  SEQUENCE FROM N.A.
RC  STRAIN-168;
RX  MEDLINE; 98044033.
RA  KUNST F., OGASAWARA N., MOSZER I., ALBERTINI A.M., ALLONI G.,
RA  AZEVEDO V., BERTERO M.G., BESSIERES P., BOLOTIN A., BORCHERT S.,
RA  BOURDIS R., BOURSIER L., BRANS A., BRAUN M., BRIGNELL S.C., BRON S.,
RA  BROUILLET S., BRUSCHI C.V., CALDWELL B., CAPUANO V., CARTER N.M.,
RA  CHOI S.K., CODANI J.J., CONNERTON I.F., CUMMINGS N.J., DANIEL R.A.,
RA  DENIZOT F., DEVINE K.M., DUSTERHOFT A., EHRLICH S.D., EMMERSON P.T.,
RA  ENTIAN K.D., ERRINGTON J., FABRET C., FERRARI E., FOULGER D.,
RA  FRITZ C., FUJITA M., FUJITA Y., FUMA S., GALIZZI A., GALLERON N.,
RA  GHM S.Y., GLASER P., GOFEAU A., GOLIGHTLY E.J., GRANDI G.,
RA  GUISEPPI G., GUY B.J., HAGA K., HALECH J., HARWOOD C.R., HENAUT A.,
RA  HILBERT H., HOLSAPPEL S., HOSONO S., HULLO M.F., IYATA M., JONES L.,
RA  JORIS B., KARAMATA D., KASAHARA Y., KLAER-BLANCHARD M., KLEIN C.,
RA  KOBAYASHI Y., KOETTER P., KORINGSTEIN G., KROGH S., KUMANO M.,
RA  KURITA K., LAPIDUS A., LARDINOIS S., LAUBER J., LAZAREVIC V.,
RA  LEE S.M., LEVINE A., LIU H., MASUDA S., MAUEL C., MEDIGUE C.,
RA  MEDINA N., MELLADO R.P., MIZUNO M., MOESTL D., NAKAI S., NOBACK M.,
RA  NOONE D., O'REILLY M., OGAWA K., OGWARA A., OUDEGA B., PARK S.H.,
RA  PARRO V., POHL T.M., PORTELELLA D., PORMOLLIK S., PRESCOTT A.M.,
RA  PRESCAN E., PUJIC P., PURNELLE B., RAPOPORT G., REY M., REYNOLDS S.,
RA  RIEGER M., RIVOLTA C., ROCHA E., ROCHE B., ROSE M., SADAIE Y.,
RA  SATO T., SCANLAN E., SCHLEICH S., SCHROETER R., SCOFFONE F.,
RA  SEKIGUCHI J., SEKORSKA A., SERO S.J., SERRO P., SHIN B.S., SOLDI B.,
RA  SROKIN A., TACCONI E., TAKAGI T., TAKAHASHI H., TAKEMARU K.,
RA  TAKEUCHI M., TAKAKOSHI A., TANAKA T., TERPSTRA P., TOGNONI A.,
RA  TOSATO V., UCHIYAMA S., VANDENBOL M., VANNIER F., VASSAROTTI A.,
RA  VIARI A., WAMBUTT R., WEDLER E., WEDLER H., WEITZENEGGER T.,
RA  WINTERS P., WIPAT A., YAMAMOTO H., YAMANE K., YASUMOTO K., YATA K.,
RA  YOSHIDA K., YOSHIKAWA H.F., ZUMSTEIN E., YOSHIKAWA H., DANCHIN A.;
RT  "The complete genome sequence of the gram-positive bacterium Bacillus
RT  subtilis."
RL  Nature 390:249-256(1997).
RN  [4]
RP  SEQUENCE FROM N.A.
RC  STRAIN-168;
RX  MEDLINE;
RA  KUNST F., OGASAWARA N., YOSHIKAWA H., DANCHIN A.;
RL  Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
DR  EMBL; AB006738; BAA21901.1;
DR  EMBL; Z99121; CAB15381.1;
SQ  SEQUENCE 323 AA; 37450 MW; 6E424868 CRC32;

Query Match 71.4%; Score 55; DB 2; Length 323;
Best Local Similarity 66.7%; Pred. No. 6.47e+00;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 71 INFEIIRYL 79
Qy 1 IPFFIVIRYL 9
[| | | | |]

RESULT 7
ID Q10836 PRELIMINARY; PRT; 1025 AA.
AC Q10836;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE THYROTROPIN-RELEASING HORMONE DEGRADING ENZYME.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN [1]

Query Match 71.4%; Score 55; DB 11; Length 1025;
Best Local Similarity 66.7%; Pred. No. 6.47e+00;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 738 IPLEIIRYL 746
Qy 1 IPFFIVIRYL 9
[| | | | |]

RESULT 8
ID P80988 PRELIMINARY; PRT; 341 AA.
AC P80988;
DT 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE THYMIDINE KINASE (EC 2.7.1.21).
GN 36
OS Human herpesvirus 3.
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Varicellovirus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 93124537.
RA TALARICO C.L., PHELPS W.C., BIRON K.K.;
RT "Analysis of the thymidine kinase genes from acyclovir-resistant
RT mutants of varicella-zoster virus isolated from patients with AIDS."
RL J. Virol. 67:1024-1033(1993).
CC -1- CATALYTIC ACTIVITY: ATP + THYMIDINE -> ADP + THYMIDINE 5'-
CC PHOSPHATE.
DR HSP; P03176; 3VTK.
DR PFAM; PF00693; TK_Herpes; 1.

```

```

RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN-SPRAGUE-DAWLEY; TISSUE=PITUITARY;
RX MEDLINE; 95023946.
RA SCHAUDER B., SCHOMBURG L., KOEHLER J., BAUER K.;
RT "Cloning of a cDNA encoding an ectoenzyme that degrades thyrotropin-
RT releasing hormone."
RL Proc. Natl. Acad. Sci. U.S.A. 91:9534-9538(1994).
CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN.
CC -1- TISSUE SPECIFICITY: PREDOMINANTLY EXPRESSED IN BRAIN AND
CC PITUITARY. LOWER LEVELS IN LUNG AND LIVER.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M1 (ZINC METALLOPROTEASE);
CC ALSO KNOWN AS THE PEPN SUBFAMILY.
DR EMBL; X80535; CAA56675.1; -.
DR PFAM; PF01433; Peptidase_M1; 1.
DR PRINTS; PR00756; ALADIPTASE.
KW Hydrolase; Metalloprotease; Zinc; Glycoprotein;
KW Sulfatation; Transmembrane; Phosphorylation; Signal-anchor.
FT DOMAIN 1 40 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 41 61 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT DOMAIN 62 1025 EXTRACELLULAR (POTENTIAL).
FT ACT_SITE 442 442 BY SIMILARITY.
FT ACT_SITE 528 528 PROTON DONOR (POTENTIAL).
FT METAL 441 441 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 445 445 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 464 464 ZINC (CATALYTIC) (BY SIMILARITY).
FT MOD_RES 30 30 PHOSPHORYLATION (BY PKC) (POTENTIAL).
FT CARBOHYD 381 381 SULFATATION (POTENTIAL).
FT CARBOHYD 90 90 POTENTIAL.
FT CARBOHYD 161 161 POTENTIAL.
FT CARBOHYD 176 176 POTENTIAL.
FT CARBOHYD 223 223 POTENTIAL.
FT CARBOHYD 339 339 POTENTIAL.
FT CARBOHYD 606 606 POTENTIAL.
FT CARBOHYD 635 635 POTENTIAL.
FT CARBOHYD 650 650 POTENTIAL.
FT CARBOHYD 664 664 POTENTIAL.
FT CARBOHYD 685 685 POTENTIAL.
FT CARBOHYD 801 801 POTENTIAL.
FT CARBOHYD 907 907 POTENTIAL.
SQ SEQUENCE 1025 AA; 117286 MW; B5CF5031 CRC32;

Query Match 71.4%; Score 55; DB 11; Length 1025;
Best Local Similarity 66.7%; Pred. No. 6.47e+00;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 738 IPLEIIRYL 746
Qy 1 IPFFIVIRYL 9
[| | | | |]

RESULT 8
ID P80988 PRELIMINARY; PRT; 341 AA.
AC P80988;
DT 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE THYMIDINE KINASE (EC 2.7.1.21).
GN 36
OS Human herpesvirus 3.
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Varicellovirus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 93124537.
RA TALARICO C.L., PHELPS W.C., BIRON K.K.;
RT "Analysis of the thymidine kinase genes from acyclovir-resistant
RT mutants of varicella-zoster virus isolated from patients with AIDS."
RL J. Virol. 67:1024-1033(1993).
CC -1- CATALYTIC ACTIVITY: ATP + THYMIDINE -> ADP + THYMIDINE 5'-
CC PHOSPHATE.
DR HSP; P03176; 3VTK.
DR PFAM; PF00693; TK_Herpes; 1.

```

```

RESULT 2
ID O77686 PRELIMINARY; PRT; 712 AA.
AC O77686;
DT 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
DE ZONA PELLUCIDA 2 PROTEIN.
GN ZP2.
OS Trichosurus vulpecula (Brush-tailed possum).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Metatheria; Diprotodontia; Phalangeridae; Trichosurus.
RN [1]
RP SEQUENCE FROM N.A.
RA MATE K.E., MCCARTNEY C.A.;
RT "Sequence and analysis of zona pellucida 2 cDNA (ZP2), from a
RT marsupial, the brushtail possum Trichosurus vulpecula.";
RL Mol. Reprod. Dev. 0:0-0(1998).
DR EMBL; AF079525; AAC28737.1; -.
DR PFAM; PF00100; zona_pellucida; 1.
DR PRINTS; PR0023; ZPELLUCIDA.
SQ SEQUENCE 712 AA; 79413 MW; 78CEB992 CRC32;

Query Match 75.3%; Score 58; DB 6; Length 712;
Best Local Similarity 85.7%; Pred. No. 1.67e+00;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 510 YPIVRYL 516
QY 3 FPIVRYL 9

RESULT 3
ID O77810 PRELIMINARY; PRT; 516 AA.
AC O77810;
DT 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
DE CYTOCHROME P-450.
GN CYP1A2.
OS Callithrix jacchus (Common marmoset).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Platyrrhini; Callitrichidae; Callithrix.
RN [1]
RP SEQUENCE FROM N.A.
RA MARMOSET CYP1A2: primary structure and constitutive expression in
RA livers.";
RL Carcinogenesis 18:1985-1991(1997).
DR EMBL; D86475; BAA33790.1; -.
DR PFAM; PF00067; P450; 1.
DR PRINTS; PR00385; P450.
DR PRINTS; PR00463; EP450I.
SQ SEQUENCE 516 AA; 58407 MW; CEF4221D CRC32;

Query Match 72.7%; Score 56; DB 6; Length 516;
Best Local Similarity 85.7%; Pred. No. 4.14e+00;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 239 FPIVRYL 245
QY 3 FPIVRYL 9

RESULT 4
ID O77809 PRELIMINARY; PRT; 516 AA.
AC O77809;
DT 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
DE CYTOCHROME P-450.

GN CYP1A2.
OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Cercopithecidae; Cercopithecinae;
OC Macaca.
RN [1]
RP SEQUENCE FROM N.A.
RA SAKUMA T., HIEDA M., OHGIYA S., NAGATA R., KAMATAKI T.;
RT "Molecular cloning and functional analysis of cynomolgus monkey
RT CYP1A2.";
RL Biochem. Pharmacol. 56:131-139(1998).
DR EMBL; D86474; BAA33789.1; -.
DR PFAM; PF00067; P450; 1.
DR PRINTS; PR00385; P450.
DR PRINTS; PR00463; EP450I.
SQ SEQUENCE 516 AA; 58197 MW; 880BAA2E CRC32;

Query Match 72.7%; Score 56; DB 6; Length 516;
Best Local Similarity 85.7%; Pred. No. 4.14e+00;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 239 FPIVRYL 245
QY 3 FPIVRYL 9

RESULT 5
ID P92553 PRELIMINARY; PRT; 145 AA.
AC P92553;
DT 01-MAY-1997 (TREMBlrel. 03, Created)
DT 01-MAY-1997 (TREMBlrel. 03, Last sequence update)
DT 01-MAY-1997 (TREMBlrel. 03, Last annotation update)
DE ORE145B.
OS Arabidopsis thaliana (Mouse-ear cress).
OG Mitochondrion.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC eudicotyledons; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
OC Arabidopsis.
RN [1]
RP SEQUENCE FROM N.A.
RA UNSELD M., MARIENFELD J.R., BRANDT P., BRENNICKE A.;
RL Nat. Genet. 0:0-0(0).
DR EMBL; Y08502; CAA69806.1; -.
KW Mitochondrion.
SQ SEQUENCE 145 AA; 17046 MW; 47CDCE7F CRC32;

Query Match 71.4%; Score 55; DB 8; Length 145;
Best Local Similarity 66.7%; Pred. No. 6.47e+00;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 132 IPFSEFVRYL 140
QY 1 IPFPIVRYL 9

RESULT 6
ID O34616 PRELIMINARY; PRT; 323 AA.
AC O34616;
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)
DE ORE1.
GN YVAX.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-JH642;
RA NAKAMURA A., GRAU R., PEREGO M., HOCH J.A.;
```

\*\*\*\*\*

W P S R E L F (TM)

\*\*\*\*\*

Release 3.1A John F. Collins, BioComputing Research Unit.

Copyright (c) 1993-1998 University of Edinburgh, U.K.

Distribution rights by Oxford Molecular Ltd

MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Apr 14 23:56:55 2000; Maspar time 15.41 Seconds

Tabular output not generated. 40.498 Million cell updates/sec

Title: >US-08-452-843-10

Description: (1-9) from US08452843.pep

Sequence: 1 IPFPVRYL 9

Scoring table: PAM 150

Gap 15

Searched: 225878 seqs, 69334122 residues

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database: sptrembl12

1:sp\_archaea 2:sp\_bacteria 3:sp\_fungi 4:sp\_human

5:sp\_invertebrate 6:sp\_mammal 7:sp\_mhc 8:sp\_organelle

9:sp\_phase 10:sp\_plant 11:sp\_rodent 12:sp\_unclassified

13:sp\_vertebrate 14:sp\_virus

Statistics: Mean 24.924; Variance 33.019; scale 0.755

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	59	76.6	1645	5 Q27448	GLUTAMINE-DEPENDENT CA	1.06e+00
2	58	75.3	712	6 Q77686	ZONA PELLUCIDA 2 PROTE	1.67e+00
3	56	72.7	516	6 Q77810	CYTOSCHROME P-450.	4.14e+00
4	56	72.7	516	6 Q77809	CYTOSCHROME P-450.	4.14e+00
5	55	71.4	145	8 P92553	ORF145B.	6.47e+00
6	55	71.4	323	2 Q34616	ORF6.	6.47e+00
7	55	71.4	1025	11 Q10836	THYROTROPIN-RELEASING	6.47e+00
8	54	70.1	341	14 P80988	THYMIDINE KINASE (EC 2	1.00e+01
9	54	70.1	492	5 Q09998	PUTATIVE 55.5 KD ZINC	1.00e+01
10	53	68.8	189	5 Q17467	B0284.3 PROTEIN.	1.55e+01
11	53	68.8	192	2 Q92366	ANTHRANILIC ACID	1.55e+01
12	53	68.8	200	2 Q66494	HYPOTHETICAL 23.0 KD P	1.55e+01
13	53	68.8	588	5 P92010	R10D12.8 PROTEIN.	1.55e+01
14	53	68.8	695	11 Q34767	ZONA PELLUCIDA 2 GLYCO	1.55e+01
15	53	68.8	948	5 Q76151	AMINOPEPTIDASE N.	1.55e+01
16	52	67.5	112	14 Q84619	GENOME, PARTIAL SEQUEN	2.38e+01
17	52	67.5	291	10 P93740	HYPOTHETICAL 32.4 KD P	2.38e+01
18	52	67.5	469	3 Q59710	FERREDOXIN-NADP+ REDUC	2.38e+01
19	52	67.5	502	2 QX231	LYSYL-TRNA SYNTHETASE.	2.38e+01
20	52	67.5	1932	5 Q01483	COSMID C06A5.	2.38e+01

21	51	66.2	170	14	O55258	THYMIDINE KINASE.	3.64e+01
22	51	66.2	210	5	Q18053	CODED FOR BY C. ELEGAN	3.64e+01
23	51	66.2	231	14	P80985	THYMIDINE KINASE (EC 2	3.64e+01
24	51	66.2	235	10	O81719	HYPOTHETICAL 26.3 KD P	3.64e+01
25	51	66.2	295	5	Q93291	C27D8.4 PROTEIN.	3.64e+01
26	51	66.2	328	14	Q96697	THYMIDINE KINASE.	3.64e+01
27	51	66.2	328	14	Q90020	THYMIDINE KINASE (EC 2	3.64e+01
28	51	66.2	341	1	O59469	341AA LONG HYPOTHETICA	3.64e+01
29	51	66.2	341	14	P80991	THYMIDINE KINASE (EC 2	3.64e+01
30	51	66.2	341	14	P80990	THYMIDINE KINASE (EC 2	3.64e+01
31	51	66.2	341	14	P80993	THYMIDINE KINASE (EC 2	3.64e+01
32	51	66.2	341	14	O57298	THYMIDINE KINASE.	3.64e+01
33	51	66.2	374	13	Q93412	ORPHAN G PROTEIN-COUP	3.64e+01
34	51	66.2	485	5	O16873	C13A2.5 PROTEIN.	3.64e+01
35	51	66.2	498	5	O17676	C49A1.5 PROTEIN.	3.64e+01
36	51	66.2	579	2	Q927V7	OLIGOPEPTIDE PERMEASE.	3.64e+01
37	51	66.2	592	11	Q64571	STEROL ESTERASE (EC 3.	3.64e+01
38	50	64.9	252	1	O58345	252AA LONG HYPOTHETICA	5.52e+01
39	50	64.9	353	2	Q9W296	HYPOTHETICAL 41.6 KD P	5.52e+01
40	50	64.9	382	4	Q13044	IONIZING RADIATION RES	5.52e+01
41	50	64.9	425	2	O67682	MODULATION COMPETITIVE	5.52e+01
42	50	64.9	555	10	O04693	GLOSSY1 HOMOLOG (FRAGM	5.52e+01
43	50	64.9	698	1	P95865	ORF C06018.	5.52e+01
44	50	64.9	845	14	Q82732	RNA-DEPENDENT RNA POLY	5.52e+01
45	50	64.9	845	14	Q9YU15	VPI PROTEIN.	5.52e+01

## ALIGNMENTS

RESULT 1  
ID Q27448 PRELIMINARY; PRT; 1645 AA.  
AC Q27448;  
DT 01-NOV-1996 (Tremblrel. 01, Created)  
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)  
DT 01-NOV-1999 (Tremblrel. 12, Last annotation update)  
DE GLUTAMINE-DEPENDENT CARBAMOYL PHOSPHATE SYNTHASE (EC 6.3.5.5)  
DE (CARBAMOYL-PHOSPHATE SYNTHASE (GLUTAMINE-HYDROLYSING))  
DE (CARBAMOYL-PHOSPHATE SYNTHETASE (GLUTAMINE-HYDROLYSING))  
DE (GLUTAMINE-DEPENDENT CARBAMOYL-PHOSPHATE SYNTHASE) (GD-CPSASE).  
GN CPSII.  
OS Babesia bovis.  
OC Eukaryota; Alveolata; Apicomplexa; Piroplasmida; Babesiidae; Babesia.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-LISMORE;  
RX MEDLINE; 96360483.  
RA BARNARA A.S.; CHANSIRI K.;  
RT "The structural gene for carbamoyl phosphate synthetase from the  
RT protozoan parasite Babesia bovis";  
RL Mol. Biochem. Parasitol. 74:239-244(1995).  
CC -1- CATALYTIC ACTIVITY: 2 ATP + GLUTAMINE + CO(2) + H(2)O = 2 ADP +  
CC PHOSPHATE + GLUTAMATE + CARBAMOYL PHOSPHATE.  
DR EMBL; U18792; AAC47302.1;  
DR HSSP; P00968; LJDB.  
DR PROSITE; PS00442; GATASE\_TYPE\_1; 1.  
DR PFAM; PF00988; CPSase\_sm\_chain; 1.  
DR PFAM; PF00117; GATase; 1.  
DR PFAM; PF00289; CPSase\_L\_chain; 3.  
DR PRINTS; PR00098; CPSASE;  
DR PRINTS; PR00099; CPSGATASE.  
DR PRINTS; PR00096; GATASE.  
KW Ligase.  
SQ SEQUENCE 1645 AA; 181549 MW; EDDD2D8D CRC32;

Query Match 76.6%; Score 59; DB 5; Length 1645;

Best Local Similarity 77.8%; Pred.No. 1.06e+01; Indels 0; Gaps 0;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 780 IAFKIVRYL 788

Qy 1 IPFPVRYL 9



**THIS PAGE BLANK (USPTO)**

REFERENCE S66676  
#authors Volz, A.; Goeke, R.; Lankat-Buttgereit, B.; Fehmann, H.C.;  
Bode, H.P.; Goeke, B.  
#journal FEBS Lett. (1995) 373:23-29  
#title Molecular cloning, functional expression, and signal  
transduction of the GIP-receptor cloned from a human  
insulinoma.  
#accession S66676  
##status preliminary  
##molecule\_type mRNA  
##residues 1-466 #label VOL  
##cross-references GB:S79852  
##note the authors translated the codon GCC for residue 427 as  
Leu  
CLASSIFICATION #superfamily glucagon receptor  
FEATURE  
1-21 #domain signal sequence #status predicted #label SIG\  
22-466 #product glucose-dependent insulinotropic protein  
receptor #status predicted #label MAT  
SUMMARY #length 466 #molecular-weight 53142 #checksum 1170  
Query Match 71.4%; Score 55; DB 2; Length 466;  
Best Local Similarity 77.8%; Pred. No. 6.85e+00;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
Db 272 IPWVIVRYL 280  
QY 1 IPFPIVRYL 9

Search completed: Fri Apr 14 23:55:33 2000  
Job time : 7 secs.

```
ACCESSIONS A33375
REFERENCE A33375
#authors Scott, S.D.; Ross, N.L.J.; Binns, M.M.
#journal J. Gen. Virol. (1989) 70:3055-3065
#title Nucleotide and predicted amino acid sequences of the Marek's
disease virus and turkey herpesvirus thymidine kinase
genes; comparison with thymidine kinase genes of other
herpesviruses.
#cross-references MUID:90063552
#accession A33375
#molecule_type DNA
#residues 1-350 #label SCO
#cross-references EMBL:D00561
CLASSIFICATION #superfamily herpesvirus thymidine kinase; herpesvirus
thymidine kinase homology
KEYWORDS ATP; DNA biosynthesis; P-loop; phosphotransferase
FEATURE 10-301 #domain herpesvirus thymidine kinase homology #label
HTK
17-24 #region nucleotide-binding motif A (P-loop)\
117-121 #region nucleotide-binding motif B\
23 #binding_site ATP (Lys) #status predicted
SUMMARY #length 350 #molecular-weight 39968 #checksum 1250

Query Match 71.4%; Score 55; DB 1; Length 350;
Best Local Similarity 85.7%; Pred. No. 6.85e+00;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 131 FPIVRYL 137
||| |||
QY 3 FPIVRYL 9

RESULT 12
ENTRY I53273 #type complete
TITLE gastric inhibitory polypeptide receptor - rat
ALTERNATE_NAMES #formal_name Rattus norvegicus #common_name Norway rat
ORGANISM 02-Aug-1996 #sequence_revision 02-Aug-1996 #text_change
DATE 04-Sep-1998
ACCESSIONS I53273
REFERENCE I53273
#authors Bonner, T.B.; Mezey, E.; Button, D.C.; Brownstein, M.J.;
Bonner, T.I.
#journal Endocrinology (1993) 133:2861-2870
#title Gastric inhibitory polypeptide receptor, a member of the
secretin-vasoactive intestinal peptide receptor family, is
widely distributed in peripheral organs and the brain.
#cross-references MUID:94062667
#accession I53273
#status preliminary; translated from GB/EMBL/DBDJ
#molecule_type mRNA
#residues 1-455 #label RES
#cross-references GB:L19660; NID:9431448; PID:9431449
CLASSIFICATION #superfamily glucagon receptor
SUMMARY #length 455 #molecular-weight 52256 #checksum 5112

Query Match 71.4%; Score 55; DB 2; Length 455;
Best Local Similarity 77.8%; Pred. No. 6.85e+00;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 269 IPWIVRYL 277
||| |||
QY 1 IPPFIVRYL 9

RESULT 13
ENTRY JC2462 #type complete
TITLE gastric inhibitory polypeptide receptor - hamster
ALTERNATE_NAMES GIP receptor
ORGANISM #formal_name Cricetinae gen. sp. #common_name hamster
DATE 15-Feb-1995 #sequence_revision 05-Apr-1995 #text_change
17-Mar-1999
ACCESSIONS JC2462
```

```
REFERENCE JC2462
#authors Yasuda, K.; Inagaki, N.; Yamada, Y.; Kubota, A.; Seino, S.;
Seino, Y.
#journal Biochem. Biophys. Res. Commun. (1994) 205:1556-1562
#title Hamster gastric inhibitory polypeptide receptor expressed in
pancreatic islets and clonal insulin-secreting cells: Its
structure and functional properties.
#cross-references MUID:95110292
#accession JC2462
#molecule_type mRNA
#residues 1-462 #label YAS
#cross-references DBJ:D38103; NID:9644880; PID:d1007862; PID:g765087
CLASSIFICATION #superfamily glucagon receptor
KEYWORDS receptor; transmembrane protein
FEATURE 136-157 #domain transmembrane #status predicted #label TM1\
167-186 #domain transmembrane #status predicted #label TM2\
215-238 #domain transmembrane #status predicted #label TM3\
252-274 #domain transmembrane #status predicted #label TM4\
292-315 #domain transmembrane #status predicted #label TM5\
339-357 #domain transmembrane #status predicted #label TM6\
383-394 #domain transmembrane #status predicted #label TM7
SUMMARY #length 462 #molecular-weight 52918 #checksum 9727

Query Match 71.4%; Score 55; DB 2; Length 462;
Best Local Similarity 77.8%; Pred. No. 6.85e+00;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 269 IPWIVRYL 277
||| |||
QY 1 IPPFIVRYL 9

RESULT 14
ENTRY G02234 #type complete
TITLE gastric inhibitory polypeptide receptor - human
ALTERNATE_NAMES GIP receptor
ORGANISM #formal_name Homo sapiens #common_name man
DATE 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change
04-Sep-1998
ACCESSIONS G02234
REFERENCE G02234
#authors Bonner, T.I.; Usdin, T.B.
#submission submitted to the EMBL Data Library, October 1995
#accession G02234
#status preliminary; translated from GB/EMBL/DBDJ
#molecule_type mRNA
#residues 1-466 #label BON
#cross-references EMBL:U39231; NID:91066050; PID:g1066051
GENETICS
#gene GDB:GIPR
#cross-references GDB:335023
#map_position 19q13.3-19q13.3
CLASSIFICATION #superfamily glucagon receptor
SUMMARY #length 466 #molecular-weight 53156 #checksum 265

Query Match 71.4%; Score 55; DB 2; Length 466;
Best Local Similarity 77.8%; Pred. No. 6.85e+00;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 272 IPWIVRYL 280
||| |||
QY 1 IPPFIVRYL 9

RESULT 15
ENTRY S66676 #type complete
TITLE glucose-dependent insulinotropic protein receptor precursor -
human
ORGANISM #formal_name Homo sapiens #common_name man
DATE 15-Feb-1997 #sequence_revision 13-Mar-1997 #text_change
04-Sep-1998
ACCESSIONS S66676
```

```

TITLE      hypothetical protein b1798 - Escherichia coli (strain K-12)
ORGANISM   #formal_name Escherichia coli
DATE       12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change
18-Sep-1998

ACCESSIONS F64940
REFERENCE   A64720
#authors   Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.;
            Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.;
            Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.;
            Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao,
            Y.

#journal   Science (1997) 277:1453-1462
#title     The complete genome sequence of Escherichia coli K-12.
#cross-references MUID:97426617
#accession F64940
#status    preliminary; nucleic acid sequence not shown;
            translation not shown

##molecule_type DNA
##residues 1-212 ##label BLAT
##cross-references GB:AE000274; GB:U00096; NID:g1788089; PID:g1788099;
            UMG:b1798

##experimental_source strain K-12, substrain MG1655
CLASSIFICATION #superfamily hypothetical protein b1798
SUMMARY        #length 212 #molecular-weight 23200 #checksum 2508

Query Match 71.4%; Score 55; DB 2; Length 212;
Best Local Similarity 77.8%; Pred. No. 6.85e+00;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 74 ILFNIVRYL 82
| | | | |
QY 1 IPFPVRYL 9

RESULT 9
ENTRY KIBETH #type complete
TITLE thymidine kinase (EC 2.7.1.21) - turkey herpesvirus
ORGANISM #formal_name turkey herpesvirus
DATE 31-Mar-1990 #sequence_revision 31-Mar-1990 #text_change
05-Sep-1997

ACCESSIONS A33346
REFERENCE Martin, S.L.; Aparicio, D.I.; Bandypadhyay, P.K.
#authors J. Virol. (1989) 63:2847-2852
#journal Genetic and biochemical characterization of the thymidine
#title kinase gene from herpesvirus of turkeys.
#cross-references MUID:89259069
#accession A33346
##molecule_type DNA
##residues 1-310 ##label MAR
##cross-references GB:M26559; NID:g330940; PID:g330941
CLASSIFICATION #superfamily herpesvirus thymidine kinase; herpesvirus
            thymidine kinase homology
KEYWORDS ATP; DNA biosynthesis; P-loop; phosphotransferase
FEATURE 10-301
            #domain herpesvirus thymidine kinase homology #label
            HTK\
17-24 #region nucleotide-binding motif A (P-loop)\
117-121 #region nucleotide-binding motif B\
23 #binding_site ATP (Lys) #status predicted
SUMMARY #length 310 #molecular-weight 35512 #checksum 7680

Query Match 71.4%; Score 55; DB 1; Length 310;
Best Local Similarity 85.7%; Pred. No. 6.85e+00;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 131 FPIRYL 137
| | | | |
QY 3 FPIRYL 9

RESULT 10
ENTRY A70029 #type complete

```

```

TITLE      hypothetical protein yvaX - Bacillus subtilis
ORGANISM   #formal_name Bacillus subtilis
DATE       05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change
24-Sep-1998

ACCESSIONS A70029
REFERENCE   A69580
#authors Kunat, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.;
            Alloni, G.; Azevedo, V.; Bertero, M.G.; Bessieres, P.;
            Bolotin, A.; Borchert, S.; Boriss, R.; Boursier, L.; Brans,
            A.; Braun, M.; Brignell, S.C.; Bron, S.; Brouillet, S.;
            Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
            Choi, S.K.; Codani, J.J.; Connerton, I.F.; Cummings, N.J.;
            Daniel, R.A.; Denizot, F.; Devine, K.M.; Duysterhoet, A.;
            Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.;
            Fabret, C.; Ferrari, E.; Foulger, D.; Fritz, C.; Fujita,
            M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galleron, N.; Ghim,
            S.Y.; Glaser, P.; Goffeau, A.; Golightly, E.J.; Grandi, G.;
            Guisseppi, G.; Guy, B.J.; Haga, K.; Halech, J.; Harwood,
            C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.;
            Hullo, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karamata, D.;
            Kasahara, Y.; Klaerr-Blanchard, M.; Klein, C.; Kobayashi,
            Y.; Koetter, P.; Konigstein, G.; Krogh, S.; Kumano, M.;
            Kurita, K.; Lapidus, A.; Lardinois, S.; Lauber, J.;
            Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.;
            Mauvel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno,
            M.; Moestl, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly,
            M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro,
            V.; Pohl, T.M.; Portetelle, D.; Porwollik, S.; Prescott, G.;
            A.M.; Prescan, E.; Pujic, P.; Purnelle, B.; Rapoport, G.;
            Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.;
            Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon, E.;
            Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.;
            Sekowska, A.; Seror, S.J.; Serror, P.; Shin, B.S.; Soldo,
            B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Takahashi, H.;
            Takemaru, K.; Takeuchi, M.; Tamakoshi, A.; Tanaka, T.;
            Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.;
            Vandenbol, M.; Vannier, F.; Vassarotti, A.; Viari, A.;
            Wambutt, R.; Wedler, E.; Wedler, H.; Weitzenecker, T.;
            Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto,
            K.; Yata, K.; Yoshida, K.; Yoshikawa, H.F.; Zumbstein, E.;
            Yoshikawa, H.; Danchin, A.
#journal   Nature (1997) 390:249-256
#title     The complete genome sequence of the Gram-positive bacterium
            Bacillus subtilis.
#cross-references MUID:98044033
#accession A70029
#status    preliminary; nucleic acid sequence not shown;
            translation not shown

##molecule_type DNA
##residues 1-323 ##label KUN
##cross-references GB:299121; GB:AL009126; NID:g2635827; PID:e1186064;
            #experimental_source strain 168

GENETICS
#gene yvaX
SUMMARY #length 323 #molecular-weight 37450 #checksum 5488

Query Match 71.4%; Score 55; DB 2; Length 323;
Best Local Similarity 66.7%; Pred. No. 6.85e+00;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 71 INFPIRYL 79
| | | | |
QY 1 IPFPVRYL 9

RESULT 11
ENTRY KIBEFC #type complete
TITLE thymidine kinase (EC 2.7.1.21) - turkey herpesvirus (strain
            FC-126)
ORGANISM #formal_name turkey herpesvirus
DATE 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change
28-Feb-1997

```

```

REFERENCE A44250
#authors Yun, C.H.; Hammons, G.J.; Jones, G.; Martin, M.V.; Hopkins,
N.E.; Alworth, W.L.; Guengerich, F.P.
#journal Biochemistry (1992) 31:10556-10563
#title Modification of cytochrome P450 1A2 enzymes by the
mechanism-based inactivator 2-ethynylnaphthalene and the
photoaffinity label 4-azidobiphenyl.
#cross-references MUID:93041749
#accession A44250
#molecule_type protein
#residues 176-185 #label YUN
#note only this tryptic peptide was photoaffinify labeled by
4-azidobiphenyl, a substrate analog
COMMENT There are three forms of this protein that differ only in the
absence or presence of the first one or two residues.
GENETICS
#gene CYP1A2
CLASSIFICATION #superfamily human cytochrome P450 CYP2D6; cytochrome P450
homology
KEYWORDS chromoprotein; electron transfer; endoplasmic reticulum;
heme; iron; monooxygenase; oxidoreductase; transmembrane
protein
FEATURE
458 #binding_site heme iron (Cys) (axial ligand) #status
predicted
SUMMARY #length 516 #molecular-weight 58334 #checksum 3338
Query Match 72.7%; Score 56; DB 1; Length 516;
Best Local Similarity 85.7%; Pred. No. 4.52e+00;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Db 239 FPIRLYL 245
III:III
QY 3 FPIVRYL 9

RESULT 5
ENTRY C37222 #type fragment
TITLE cytochrome P450 1A1, hepatic - dog (fragment)
ALTERNATE_NAMES cytochrome P450 (Dahl)
CONTAINS oxidoreductase (EC 1.-.-)
ORGANISM #formal_name Canis lupus familiaris #common_name dog
DATE 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change
05-Mar-1999
ACCESSIONS C37222
REFERENCE A37222
#authors Uchida, T.; Komori, M.; Kitada, M.; Kamataki, T.
#journal Mol. Pharmacol. (1990) 38:644-651
#title Isolation of cDNAs coding for three different forms of liver
microsomal cytochrome P-450 from polychlorinated
biphenyl-treated beagle dogs.
#cross-references MUID:91042464
#accession C37222
#status not compared with conceptual translation
#molecule_type mRNA
#residues 1-524 #label UCH
CLASSIFICATION #superfamily human cytochrome P450 CYP2D6; cytochrome P450
homology
KEYWORDS chromoprotein; electron transfer; heme; iron; liver;
monooxygenase; oxidoreductase; transmembrane protein
FEATURE
461 #binding_site heme iron (Cys) (axial ligand) #status
predicted
SUMMARY #length 524 #checksum 6556
Query Match 72.7%; Score 56; DB 2; Length 524;
Best Local Similarity 85.7%; Pred. No. 4.52e+00;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Db 241 FPIRLYL 247
III:III
QY 3 FPIVRYL 9

REFERENCE A44250
#authors Yun, C.H.; Hammons, G.J.; Jones, G.; Martin, M.V.; Hopkins,
N.E.; Alworth, W.L.; Guengerich, F.P.
#journal Biochemistry (1992) 31:10556-10563
#title Modification of cytochrome P450 1A2 enzymes by the
mechanism-based inactivator 2-ethynylnaphthalene and the
photoaffinity label 4-azidobiphenyl.
#cross-references MUID:93041749
#accession A44250
#molecule_type protein
#residues 176-185 #label YUN
#note only this tryptic peptide was photoaffinify labeled by
4-azidobiphenyl, a substrate analog
COMMENT There are three forms of this protein that differ only in the
absence or presence of the first one or two residues.
GENETICS
#gene CYP1A2
CLASSIFICATION #superfamily human cytochrome P450 CYP2D6; cytochrome P450
homology
KEYWORDS chromoprotein; electron transfer; endoplasmic reticulum;
heme; iron; monooxygenase; oxidoreductase; transmembrane
protein
FEATURE
458 #binding_site heme iron (Cys) (axial ligand) #status
predicted
SUMMARY #length 516 #molecular-weight 58334 #checksum 3338
Query Match 72.7%; Score 56; DB 1; Length 516;
Best Local Similarity 85.7%; Pred. No. 4.52e+00;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Db 239 FPIRLYL 245
III:III
QY 3 FPIVRYL 9

RESULT 6
ENTRY S70397 #type complete
TITLE zona pellucida glycoprotein A - dog
ORGANISM #formal_name Canis lupus familiaris #common_name dog
DATE 28-Oct-1996 #sequence_revision 27-Feb-1997 #text_change
17-Mar-1999
ACCESSIONS S70397
REFERENCE S70396
#authors Harris, J.D.; Hibler, D.W.; Fontenot, G.K.; Hsu, K.T.;
Yurewicz, E.C.; Sacco, A.G.
#journal DNA Seq. (1994) 4:361-393
#title Cloning and characterization of zona pellucida genes and
cDNAs from a variety of mammalian species: the ZPA, ZPB and
ZPC gene families.
#cross-references MUID:95143578
#accession S70397
#status preliminary
#molecule_type mRNA
#residues 1-715 #label HAR
#cross-references EMBL:U05779; NID:G458274; PID:G458275
CLASSIFICATION #superfamily sperm-binding glycoprotein ZP2; ZP domain
homology
FEATURE
368-628 #domain ZP domain homology #label ZPH
SUMMARY #length 715 #molecular-weight 79938 #checksum 3009
Query Match 72.7%; Score 56; DB 2; Length 715;
Best Local Similarity 71.4%; Pred. No. 4.52e+00;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
Db 511 YPVRYL 517
III:III
QY 3 FPIVRYL 9

RESULT 7
ENTRY S70398 #type complete
TITLE zona pellucida glycoprotein A - cat
ORGANISM #formal_name Felis silvestris catus #common_name domestic cat
DATE 28-Oct-1996 #sequence_revision 27-Feb-1997 #text_change
17-Mar-1999
ACCESSIONS S70398
REFERENCE S70396
#authors Harris, J.D.; Hibler, D.W.; Fontenot, G.K.; Hsu, K.T.;
Yurewicz, E.C.; Sacco, A.G.
#journal DNA Seq. (1994) 4:361-393
#title Cloning and characterization of zona pellucida genes and
cDNAs from a variety of mammalian species: the ZPA, ZPB and
ZPC gene families.
#cross-references MUID:95143578
#accession S70398
#status preliminary
#molecule_type mRNA
#residues 1-716 #label HAR
#cross-references EMBL:U05776; NID:G458268; PID:G458269
CLASSIFICATION #superfamily sperm-binding glycoprotein ZP2; ZP domain
homology
FEATURE
370-630 #domain ZP domain homology #label ZPH
SUMMARY #length 716 #molecular-weight 80135 #checksum 6483
Query Match 72.7%; Score 56; DB 2; Length 716;
Best Local Similarity 71.4%; Pred. No. 4.52e+00;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
Db 513 YPVRYL 519
III:III
QY 3 FPIVRYL 9

RESULT 8
ENTRY F64940 #type complete

```

```

REFERENCE
#authors A90953 PID:g181317
#journal Quattrochi, L.C.; Okino, S.T.; Pendurthi, U.R.; Tukey, R.H.
#title DNA (1985) 4:395-400
#cross-references MUID:86081170
#accession A23585
#molecule_type mRNA
#residues 295-310,'L',312-449,'W',450,'V',452-485 ##label Q02
#cross-references GB:M12078; NID:g181351; PID:g553246
REFERENCE
#authors S07373
#journal Jaiswal, A.K.; Nebert, D.W.; Gonzalez, F.J.
#title Human P(3)450: cDNA and complete amino acid sequence.
#cross-references MUID:86312938
#accession S07373
#status translation not shown
#molecule_type mRNA
#residues 1-515 ##label JA1
#cross-references EMBL:200036; NID:g30338; PID:g30339
REFERENCE
#authors S22433
#journal Jaiswal, A.K.; Nebert, D.W.; McBride, O.W.; Gonzalez, F.J.
#title Human P(3)450: cDNA and complete protein sequence, repetitive
#cross-references MUID:88061719
#accession S22433
#status preliminary
#molecule_type mRNA
#residues 1-515 ##label JA2
#cross-references EMBL:W5053; NID:g181307; PID:g181308
REFERENCE
#authors A60881
#journal Wrighton, S.A.; Campanile, C.; Thomas, P.E.; Maines, S.L.;
#title Shively, J.E.; Levin, W.; Guzellan, P.S.
Mol. Pharmacol. (1986) 29:405-410
Identification of a human liver cytochrome P-450 homologous
to the major isosafrole-inducible cytochrome P-450 in the
rat.
#cross-references MUID:86203234
#accession A60881
#molecule_type protein
#residues 2-19 ##label WRI
GENETICS
#gene GDB:CYP1A2
#cross-references GDB:118780; OMIM:124060
#map_position 15q22-15qter
#introns 277/3; 318/1; 348/1; 389/2; 418/2
CLASSIFICATION
#superfamily human cytochrome P450 CYP2D6; cytochrome P450
homology
KEYWORDS
chromoprotein; electron transfer; endoplasmic reticulum;
heme; iron; microsome; monooxygenase; oxidoreductase;
transmembrane protein
FEATURE
458 #binding_site heme iron (Cys) (axial ligand) #status
predicted
SUMMARY
#length 515 #molecular_weight 58294 #checksum 960
Query Match 72.7%; Score 56; DB 1; Length 515;
Best Local Similarity 85.7%; Pred. No. 4.52e+00;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Db 239 PFILRYL 245
Oy 3 PFIVRYL 9
RESULT 4
ENTRY #type complete
TITLE cytochrome P450 1A2 - rabbit

```

---

```

ALTERNATE_NAMES acetanilide 4-hydroxylase (EC 1.14.14.-); cytochrome P450
LM4; cytochrome P450-4
ORGANISM #formal_name Oryctolagus cuniculus #common_name domestic
rabbit
DATE 28-May-1986 #sequence_revision 24-Feb-1994 #text_change
ACCESSIONS B27821; S02038; B25143; A00187; A00188; A44250
REFERENCE #authors Kagawa, N.; Mihara, K.; Sato, R.
#journal J. Biochem. (1987) 101:1471-1479
#title Structural analysis of cloned cDNAs for polycyclic
hydrocarbon-inducible forms of rabbit liver microsomal
cytochrome P-450.
#cross-references MUID:88032911
#accession B27821
#molecule_type mRNA
#residues 1-120,'H',122-516 ##label KA2
#cross-references EMBL:X05686; NID:g1540; PID:g1541
REFERENCE #authors S02038
#journal Pompon, D.
#title Eur. J. Biochem. (1988) 177:285-293
cDNA cloning and functional expression in yeast Saccharomyces
cerevisiae of beta-naphthoflavone-induced rabbit liver
P-450 LM4 and LM6.
#cross-references MUID:89052697
#accession S02038
#molecule_type mRNA
#residues 1-173,'S',175-207,'H',209-232,'S',234-298,'G',300-353,
'PG',356-516 ##label POM
#cross-references EMBL:X13853; NID:g1532; PID:g1533
#note the authors translated the codon GAC for residue 276 as
Gln and CCC for residue 354 as Ala
REFERENCE #authors A94056
#journal Okino, S.T.; Quattrochi, L.C.; Barnes, H.J.; Osanto, S.;
Griffin, K.J.; Johnson, E.F.; Tukey, R.H.
#title Proc. Natl. Acad. Sci. U.S.A. (1985) 82:5310-5314
Cloning and characterization of cDNAs encoding 2,3,7,
8-tetrachlorodibenzo-p-dioxin-inducible rabbit mRNAs for
cytochrome P-450 isozymes 4 and 6.
#cross-references MUID:85270514
#accession B25143
#molecule_type mRNA
#residues 'H',94-207,'H',209-287,'I',289-290,'N',292,'MD',295,
'MDDGAHV',303-308,'T',310-357,'L',359-461,'I',463-516
##label OKI
#cross-references EMBL:M11728; NID:g165578; PID:g165579
REFERENCE #authors A00187
#journal Ozols, J.
#title J. Biol. Chem. (1986) 261:3965-3979
Complete amino acid sequence of a cytochrome P-450 isolated
from beta-naphthoflavone-induced rabbit liver microsomes.
Comparison with phenobarbital-induced and constitutive
isozymes and identification of invariant residues.
#cross-references MUID:86140205
#accession A00187
#molecule_type protein
#residues 2-21,'S',23-69,'Q',71-91,'N',93-171,'F',173-193,'S',
195-207,'PPOGM',213-246,'OPN',250,'R',252-289,'SH',
292-294;296-298,'G',300-493,'T',495-516 ##label O20
#note 233-Ser and 247-Asn were also found
REFERENCE #authors A00188
#journal Fujita, V.S.; Black, S.D.; Tarr, G.E.; Koop, D.R.; Coon, M.J.
#title Proc. Natl. Acad. Sci. U.S.A. (1984) 81:4260-4264
On the amino acid sequence of cytochrome P-450 isozyme 4 from
rabbit liver microsomes.
#cross-references MUID:84272618
#accession A00188
#molecule_type protein
#residues 2-45,'S',47,'V',49;107-118;133-173,'S',175-197,'X',
199-206;217-232,'S',234-241,'V',243-246;255-264;
267-274;297-298,'G',300-341;362-376,'XX',379-381;
394-444,'A',446-477,'X',479-486;500-511,'S',513;'K',
##label FUJ

```

```

#cross-references MUID:98044033
#accession A69715
#status nucleic acid sequence not shown; translation not shown
#molecule_type DNA
#residues 1-492 #label KUN
#cross-references GB:Z99116; GB:AL009126; NID:G2634723; PID:el185608;
#experimental_source strain 168
#accession S45533
#authors Sorokin, A.; Zumstein, E.; Azevedo, V.; Ehrlich, S.D.;
#submission submitted to the EMBL Data Library, November 1993
#accession S45533
#molecule_type DNA
#residues 81-431, 'L', 433-492 #label SOR
#cross-references EMBL:L09228; NID:G410114; PID:G410115
#accession J00471
#authors Yamamoto, J.; Shimizu, M.; Yamane, K.
#journal Agric. Biol. Chem. (1991) 55:1615-1626
#title Molecular cloning and analysis of nucleotide sequence of the
Bacillus subtilis lysA gene region using B. subtilis phage
vectors and a multi-copy plasmid, pUB110.
#cross-references MUID:91345841
#accession PS0430
#molecule_type DNA
#residues 223-431, 'L', 433-466, 'P', 468-492 #label YAM
#note the authors translated the codon TCA for residue 392 as
Thr, ACT for residue 419 as Ser, TAT for residues 425
and 437 as Thr, CTT for residue 432 as Ile, TAC for
residue 433 as Thr, and AGC for residue 478 as Thr

GENETICS
#gene spoVAF
KEYWORDS sporulation; transmembrane protein
SUMMARY #length 492 #molecular-weight 55606 #checksum 2481

Query Match 74.0%; Score 57; DB 2; Length 492;
Best Local Similarity 85.7%; Pred. No. 2.96e+00;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 232 PFPIVRY 238
|||||
Qy 2 PFPIVRY 8

RESULT 2
ENTRY JX0190 #type complete
TITLE Cytochrome P450 1A2 - golden hamster
ALTERNATE_NAMES cytochrome P450 MC4; cytochrome P450-H (2,3,4,7,
8-pentachlorodibenzofuran inducible)
CONTAINS oxidoreductase (EC 1.-.-.-)
ORGANISM #formal_name Mesocricetus auratus #common_name golden hamster
DATE 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change
05-Mar-1999
ACCESSIONS JX0190; S13885; PX0036
REFERENCE JX0189
#authors Sagami, I.; Ohmachi, T.; Fujii, H.; Kikuchi, H.; Watanabe, M.
#journal J. Biochem. (1991) 110:641-647
#title Hamster cytochrome P-450 1A gene family, P-450 1A1 and P-450-
1A2 in lung and liver: cDNA cloning and sequence analysis.
#cross-references MUID:92138673
#accession JX0190
#molecule_type mRNA
#residues 1-513 #label SAG
#experimental_source lung and liver, microsomes
#accession S13884
#authors Lai, T.S.; Chiang, J.Y.L.
#journal Arch. Biochem. Biophys. (1990) 283:429-439
#title Cloning and characterization of two major
3-methylcholanthrene inducible hamster liver cytochrome
P450s.
#cross-references MUID:91112759
#accession S13885
#molecule_type mRNA

```

```

#residues 1-48, 'F', 50-51, 'MC', 54-252, 'GG', 255-325, 'W', 327-355, 'L',
357-484, 'Q', 486-513 #label LAI
#cross-references EMBL:M63787; NID:G191354; PID:G191355
#accession PX0036
#authors Koga, N.; Ariyoshi, N.; Nakashima, H.; Yoshimura, H.
#journal J. Biochem. (1990) 107:826-833
#title Purification and characterization of two forms of 2,3,4,7,
8-pentachlorodibenzofuran-inducible cytochrome P-450 in
hamster liver.
#cross-references MUID:90361684
#accession PX0036
#molecule_type protein
#residues 2-19 #label KOG
#experimental_source liver
#comment Cytochrome P-450 I family consists of two members, 1A1 and 1A2.
Both of them are inducible by 3-methylcholanthrene and 2,3,7,
8-tetrachlorodibenzo-p-dioxin, but have different substrate
specificities.

GENETICS
#gene CYP1A2
CLASSIFICATION #superfamily human cytochrome P450 CYP2D6; cytochrome P450
homology
KEYWORDS chromoprotein; electron transfer; endoplasmic reticulum;
heme; iron; monooxygenase; oxidoreductase; transmembrane
protein
FEATURE
456 #binding_site heme iron (Cys) (axial ligand) #status
predicted
SUMMARY #length 513 #molecular-weight 58082 #checksum 3323

Query Match 72.7%; Score 56; DB 2; Length 513;
Best Local Similarity 85.7%; Pred. No. 4.52e+00;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 238 PFPIVRYL 244
|||||
Qy 3 PFPIVRYL 9

RESULT 3
ENTRY O4HU4 #type complete
TITLE cytochrome P450 1A2 - human
ALTERNATE_NAMES cytochrome P450 HLD; cytochrome P450-4; cytochrome P450-P3
CONTAINS oxidoreductase (EC 1.-.-.-)
ORGANISM #formal_name Homo sapiens #common_name man
DATE 28-Dec-1987 #sequence_revision 31-Mar-1992 #text_change
05-Mar-1999
ACCESSIONS S16718; A25892; A23585; S07373; S22433; A60881
REFERENCE S16718
#authors Ikeya, K.; Jaiswal, A.K.; Owens, R.A.; Jones, J.E.; Nebert,
D.W.; Kimura, S.
#journal Mol. Endocrinol. (1989) 3:1399-1408
#title Human CYP1A2: sequence, gene structure, comparison with the
mouse and rat orthologous gene, and differences in liver
1A2 mRNA expression.
#cross-references MUID:90114205
#accession S16718
#molecule_type DNA
#residues 1-515 #label IKE
#cross-references EMBL:M31664; NID:G181377; EMBL:M31665; NID:G181378;
EMBL:M31666; NID:G181379; EMBL:M31667; NID:G181380;
PID:G181382
#accession A25892
#authors Quattrocchi, L.C.; Pendurthi, U.R.; Okino, S.T.; Potenza, C.;
Tukey, R.H.
#journal Proc. Natl. Acad. Sci. U.S.A. (1986) 83:6731-6735
#title Human cytochrome P-450 4 mRNA and gene: part of a multigene
family that contains Alu sequences in its mRNA.
#cross-references MUID:86313652
#accession A25892
#molecule_type DNA
#residues 1-78, 'S', 80-510, 'LP', 512-515 #label QUA
#cross-references EMBL:L00388; GB:L00389; EMBL:M14337; NID:G181315;

```

\*\*\*\*\*  
Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd  
MParch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Fri Apr 14 23:55:26 2000; MasPar time 3.64 Seconds  
Tabular output not generated.  
99.111 Million cell updates/sec  
\*\*\*\*\*

Title: >US-08-452-843-10  
Description: (1-9) from US08452843.pep  
Perfect Score: 77  
Sequence: 1 IPPPIVRYL 9  
Scoring table: PAM 150  
Gap 15  
Searched: 122810 seqs, 40068593 residues  
Post-processing: Minimum Match 0%  
Listing first 45 summaries  
Database: pir62  
1:pir1 2:pir2 3:pir3 4:pir4  
Statistics: Mean 25.168; Variance 35.578; scale 0.707

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	57	74.0	492	2 A69715	spore formation prote	2.96e+00
2	56	72.7	513	2 JX0190	cytochrome P450 1A2 -	4.52e+00
3	56	72.7	513	1 O4H04	cytochrome P450 1A2 -	4.52e+00
4	56	72.7	516	1 O4RBN	cytochrome P450 1A2 -	4.52e+00
5	56	72.7	524	2 C37222	cytochrome P450 1A1,	4.52e+00
6	56	72.7	715	2 S70397	zona pellucida glycop	4.52e+00
7	56	72.7	716	2 S70398	zona pellucida glycop	4.52e+00
8	55	71.4	212	2 F64940	hypothetical protein	6.85e+00
9	55	71.4	310	1 KIBETH	thymidine kinase (EC	6.85e+00
10	55	71.4	323	2 A70029	hypothetical protein	6.85e+00
11	55	71.4	350	1 KIBEFC	thymidine kinase (EC	6.85e+00
12	55	71.4	455	2 I53273	gastric inhibitory po	6.85e+00
13	55	71.4	462	2 JC2462	gastric inhibitory po	6.85e+00
14	55	71.4	466	2 G02234	gastric inhibitory po	6.85e+00
15	55	71.4	466	2 S66876	glucose-dependent ins	6.85e+00
16	55	71.4	491	2 I37411	glucose-dependent ins	6.85e+00
17	55	71.4	1025	2 I59331	thyrotropin-releasing	6.85e+00
18	54	70.1	513	1 O4MGM3	acetanilide 4-hydroxy	1.03e+01
19	54	70.1	513	2 A61400	cytochrome P450 1A2 -	1.03e+01
20	53	68.8	189	2 S43558	membrane protein B028	1.55e+01
21	53	68.8	200	2 C70308	hypothetical protein	1.55e+01
22	53	68.8	276	2 A45984	sperm-binding glycopr	1.55e+01
23	53	68.8	713	2 A34782	sperm-binding glycopr	1.55e+01

24	52	67.5	296	2 H65118	hypothetical adenine-	2.32e+01
25	52	67.5	1064	2 S52687	serine/threonine-spec	2.32e+01
26	51	66.2	242	2 G64305	phosphoribosylamino	3.44e+01
27	51	66.2	261	2 F64305	hypothetical protein	3.44e+01
28	51	66.2	275	2 S40005	trypsin (EC 3.4.21.4)	3.44e+01
29	51	66.2	341	1 KIBE73	thymidine kinase (EC	3.44e+01
30	51	66.2	341	2 E71191	probable 3-hydroxy-3-	3.44e+01
31	51	66.2	341	1 KIBE36	thymidine kinase (EC	3.44e+01
32	51	66.2	341	1 KIBEGK	thymidine kinase (EC	3.44e+01
33	51	66.2	341	1 KIBEEL	thymidine kinase (EC	3.44e+01
34	51	66.2	341	1 KIBE40	thymidine kinase (EC	3.44e+01
35	51	66.2	380	2 JQ2338	omega-3 fatty acid de	3.44e+01
36	51	66.2	463	2 A46172	glucagon-like peptide	3.44e+01
37	51	66.2	563	2 A70038	L-lactate permease ho	3.44e+01
38	51	66.2	612	2 A34967	sterol esterase (EC 3	3.44e+01
39	51	66.2	713	2 S70434	zona pellucida glycop	3.44e+01
40	50	64.9	227	2 S74918	biopolymer transport	5.08e+01
41	50	64.9	252	2 B71105	hypothetical protein	5.08e+01
42	50	64.9	382	2 G01589	ionizing radiation re	5.08e+01
43	50	64.9	425	2 H70456	modulation competitiv	5.08e+01
44	50	64.9	461	2 T01825	hypothetical protein	5.08e+01
45	50	64.9	845	1 RRXSJA	RNA-directed RNA poly	5.08e+01

ALIGNMENTS

RESULT 1  
ENTRY #type complete  
TITLE Spore formation protein spovAF - Bacillus subtilis  
ALTERNATE\_NAMES stage V sporulation protein AF  
ORGANISM #formal\_name Bacillus subtilis  
DATE 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 17-Mar-1999  
ACCESSIONS A69715; S45533; PS0430  
REFERENCE A69580  
#authors Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bertero, M.G.; Bessieres, P.; Bolotin, A.; Borchert, S.; Boriss, R.; Boursier, L.; Brans, A.; Braun, M.; Brignelli, S.C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Choi, S.K.; Codani, J.J.; Connerton, I.F.; Cummings, N.J.; Daniel, R.A.; Denizot, F.; Devine, K.M.; Duescherhoef, A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.; Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galleron, N.; Ghim, S.Y.; Glaser, P.; Goffeau, A.; Golightly, E.J.; Grandi, G.; Guisepi, G.; Guy, B.J.; Haga, K.; Halech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karamata, D.; Kasahara, Y.; Klaerr-Blanchard, M.; Klein, C.; Kobayashi, Y.; Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois, S.; Lauber, J.; Lazarevic, V.; Lee, S.W.; Levine, A.; Liu, H.; Masuda, S.; Maelen, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno, M.; Moestl, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle, D.; Porwollik, S.; Prescott, A.M.; Praecan, E.; Pujic, P.; Purnelle, B.; Rapoport, G.; Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon, E.; Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror, S.J.; Serror, P.; Shin, B.S.; Soldo, B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Takahashi, H.; Takemaru, K.; Takeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.; Vandenbol, R.; Vannier, F.; Vassarotti, A.; Viari, A.; Wandt, R.; Wedler, E.; Wedler, H.; Weitzenecker, T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.; Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.

Nature (1997) 390:249-256  
The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.



**THIS PAGE BLANK (USPTO)**

**THIS PAGE BLANK (USPTO)**

PR 29-JAN-1993; US-012990.  
 PR 07-JUN-1995; US-484993.  
 PA (ZONA-) ZONAGEN INC.  
 PI Harris JD;  
 DR WPI: 99-023447/02.  
 DR N-PSDB: V84792.

PT Isolated zona pellucida DNA from different mammals - used to develop  
 PT products which can be used for vaccination to induce transient  
 PT infertility or permanent sterility in female mammals  
 PS Example 4: Column 85-88; 84pp; English.  
 CC This sequence represents a feline ZPA protein isolated from zona  
 CC pellucida. This protein can be used in a method for specifically  
 CC inducing transient infertility or permanent sterility in a host  
 CC animal by selective vaccination with specific zona pellucida proteins  
 CC or immunocontraceptively active fragments.  
 SQ Sequence 716 AA;

Query Match 72.7%; Score 56; DB 1; Length 716;  
 Best Local Similarity 71.4%; Pred. No. 3.96e+01;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 513 YPVVRYL 519  
 :|:||||  
 Qy 3 FPIVRYL 9

## RESULT 13

ID R60532 standard; Protein; 716 AA.  
 AC R60532;  
 DT 12-MAY-1995 (first entry)  
 DE Feline zona pellucida (FZP)-2.  
 KW Feline zona pellucida; FZP-2; contraceptive vaccine antigen.  
 OS Felis domesticus.  
 PN J0621777-A.  
 PD 09-AUG-1994.  
 PF 29-JAN-1993; 013496.  
 PR 29-JAN-1993; JP-013496.  
 PA (TOFU ) TONEN CORP.  
 DR WPI: 94-305384/38.  
 DR N-PSDB: Q71287.  
 PT New DNA sequence coding feline zona pellucida (FZP) 2 - used for  
 PT prodn. of contraceptive vaccine antigen for cats  
 PS Claim 1; Page 7-10; 10pp; Japanese.  
 CC The feline zona pellucida (FZP)-2 coding sequence was obtained by  
 CC PCR amplification of cDNA synthesised from mRNA isolated from cat  
 CC ovaries. Subfragments of the 716 amino acid FZP-2 protein can be  
 CC expressed for use as antigens in contraceptive vaccines for cats.  
 SQ Sequence 716 AA;

Query Match 72.7%; Score 56; DB 1; Length 716;  
 Best Local Similarity 71.4%; Pred. No. 3.96e+01;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 513 YPVVRYL 519  
 :|:||||  
 Qy 3 FPIVRYL 9

## RESULT 14

ID R89369 standard; peptide; 9 AA.  
 AC R89369;  
 DT 18-SEP-1996 (first entry)  
 DE Cw6 consensus peptide derived immunogenic peptide #1.  
 KW Immunogenic peptide; supermotif; HLA molecule; CTL response;  
 KW therapeutic; diagnostic; cancer; viral infection; hepatitis B;  
 KW hepatitis C.  
 OS Synthetic.  
 PN WO9603140-A1.  
 PD 08-FEB-1996.  
 PF 21-JUL-1995; U09234.  
 PR 21-JUL-1994; US-278634.  
 PR 23-NOV-1994; US-344824.  
 PR 30-MAY-1995; US-452843.

PA (CYTE-) CYTEL CORP.  
 PI Sette A, Sidney J;  
 DR WPI: 96-116784/12.  
 PT Compn. comprising immunogenic peptide with supermotif allowing more  
 PT than one HLA mol. to bind - used to induce CTL response in patient  
 PT and for in vivo and ex vivo therapeutic and diagnostic applications  
 PS Claim 2; Page 26; 32pp; English.

CC The sequences given in R89362-82 are immunogenic peptides which were  
 CC use in the composition of the invention. The composition comprises  
 CC an immunogenic peptide of 9-10 residues with a supermotif which  
 CC allows binding of more than one HLA molecule. It pref. comprises  
 CC two conserved residues, a first at the 2nd position from the N-  
 CC terminal is Pro, and a 2nd at the C-terminal is Met. These peptides  
 CC are used to induce a CTL response in a patient. They are also  
 CC useful in compositions for in vivo and ex vivo therapeutic and  
 CC diagnostic applications, e.g. the treatment of cancer and viral  
 CC infections, e.g. hepatitis B and C.  
 SQ Sequence 9 AA;

Query Match 68.8%; Score 53; DB 1; Length 9;  
 Best Local Similarity 85.7%; Pred. No. 8.25e+01;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 1 IPYPIVR 7  
 ||:||||  
 Qy 1 IFFPIVR 7

## RESULT 15

ID R06998 standard; protein; 713 AA.  
 AC R06998;  
 DT 18-JAN-1991 (first entry)  
 DE Mouse ZP2 protein exhibiting epitope for Ab which inhibits fertilisation  
 DE of an oocyte by a sperm.  
 KW Contraceptive; ZP3 protein; zona pellucida.  
 OS Mus musculus.  
 PN US7364379-A.  
 PD 28-AUG-1990.  
 PF 12-JUN-1989; 364379.  
 PR 12-JUN-1989; US-364379.  
 PA (USSH ) NAT INST DIABETES.  
 PI Jurrien D;  
 DR WPI: 90-297734/39.  
 DR NP-PSDB: Q06005.  
 PT Contraceptive antibody vaccine for mammalian female - comprises  
 PT peptide epitope of zona pellucida protein, minimises possibility  
 PT of birth defects if failed contraception.  
 PS Disclosure; Fig 3; 93pp; English.  
 CC Vaccine provides long term, non-permanent contraception in mammals,  
 CC by inhibition of fertilisation rather than abortive methods, thus  
 CC minimising risk of birth defects.  
 CC Gene product comprises epitope to zona pellucida protein and vectors  
 CC and transformed expression systems are also claimed.  
 SQ Sequence 713 AA;

Query Match 68.8%; Score 53; DB 1; Length 713;  
 Best Local Similarity 71.4%; Pred. No. 8.25e+01;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 507 YPLVRYL 513  
 :|:||||  
 Qy 3 FPIVRYL 9

Search completed: Fri Apr 14 23:55:09 2000  
 Job time : 41 secs.



PF 15-JUL-1994; 164184.  
PR 20-JUL-1993; JP-201120.  
PR 30-JUL-1993; JP-208279.  
PR 17-JUN-1994; JP-136053.  
PA (SUMO) SUMITOMO CHEM CO LTD.  
DR WPI; 96-182311/19.  
DR N-PSDB; T28380.  
PT Novel method for the evaluation of the safety of a cpd. - using a  
PT human cytochrome P450 and yeast NADPH reductase to determine whether  
PT the analyte cpd. is detoxified or metabolised to a carcinogen  
PS Example 1; Page 18-20; 74pp; Japanese.  
CC This is the amino acid sequence of the human cytochrome P450 molecular  
CC species 1A2 protein. The corresp. 1.5 kb fragment encoding the protein  
CC was amplified from a human liver derived cDNA library using primers  
CC T26923-4. The prod. was cloned into the yeast expression vector pAAH5N  
CC to generate plasmid p1A2 for prodn. of the cytochrome only or into the  
CC vector pAHRR to generate the plasmid p1A2R for co-prodn. with the yeast  
CC NADPH-P450 reductase. The sequence is placed under control of the yeast  
CC ADH gene promoter and terminator.  
CC The vectors are used in a method for evaluating the safety of a cpd. by  
CC reacting the test cpd. with recombinantly produced human cytochrome P450  
CC mol. species 1A2, 2C9 (T28381), 2E1 (T28382), 3A4 (T28383) or their  
CC variants (T28384-98) together with yeast NADPH-P450 reductase (either as  
CC a fused protein or as a cell extract) and analysing the resultant  
CC metabolite. The cpd. is considered "safe" if it is detoxified or not  
CC rendered carcinogenic or "unsafe" if it is not detoxified or is  
CC metabolised to a carcinogenic cpd.  
SQ Sequence 516 AA;

Query Match 72.7%; Score 56; DB 1; Length 516;  
Best Local Similarity 85.7%; Pred. No. 3.96e+01;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 239 FPIVRYL 245  
QY 3 FPIVRYL 9  
|||||

RESULT 6  
ID R72360 standard; Protein; 516 AA.  
AC R72360; 1995 (first entry)  
DE Human cytochrome P450 molecular species 1A2 protein.  
KW Human cytochrome P450; amplification; PCR; primer; expression vector;  
KW yeast NADPH-P450 reductase; safety; fusion protein; metabolite;  
KW Carcinogen; mutagen; liver metabolism.  
OS Homo sapiens.  
PN EP-644267-A.  
PD 22-MAR-1995.  
PF 20-JUL-1994; 111298.  
PR 20-JUL-1993; JP-201120.  
PR 21-JUL-1993; JP-180246.  
PR 30-JUL-1993; JP-208279.  
PA (HAYA/) HAYASHI K.  
PA (SUMO) SUMITOMO CHEM CO LTD.  
PI Hayashi K, Kaneko H, Komai K, Nakatsuka I, Sakaki T;  
PI Yabusaki Y;  
DR WPI; 95-116991/16.  
DR N-PSDB; Q87114.  
PT Evaluation of safety of a chemical cpd. - using recombinant yeast  
PT expressing human cytochrome P450 and a yeast NADPH-P450 reductase  
PS Examples; Page 18-21; 124pp; English.  
CC The amino acid sequence of the human cytochrome P450 species 1A2. The  
CC 1.5 kb cDNA was amplified by PCR using the primers Q87733-4. The product  
CC was cloned into the yeast expression vectors pAAH5N or pAHRR to produce  
CC the vectors p1A2 for the expression of the cytochrome P450 alone or p1A2R  
CC co-expressed with the yeast NADPH-P450 reductase, respectively.  
CC The vectors are used in a method for evaluating the safety of a chemical  
CC compound by reacting the chemical compound with recombinantly produced  
CC human cytochrome P450 molecular species 1A2, 2C9 (Q87715), 2E1 (Q87716)  
CC or 3A4 (Q87717) or their auxiliary species and variants (Q87718-32) and  
CC yeast NADPH-P450 reductase, either as a fused protein or in cell  
CC extracts, and analysing the resulting metabolite to assess the safety of

CC the chemical compound. The method is useful for determining whether the  
CC chemical compound, or its metabolite, will be converted into a  
CC carcinogenic or mutagenic form through metabolism in the liver.  
SQ Sequence 516 AA;

Query Match 72.7%; Score 56; DB 1; Length 516;  
Best Local Similarity 85.7%; Pred. No. 3.96e+01;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 239 FPIVRYL 245  
QY 3 FPIVRYL 9  
|||||

RESULT 7  
ID W00183 standard; Protein; 516 AA.  
AC W00183; 1996 (first entry)  
DE Cytochrome P4501A2.  
KW Primer; polymerase chain reaction; PCR; amplify cytochrome P4501A2;  
KW human; antibody; detection.  
OS Synthetic.  
PN J08143600-A.  
PD 04-JUN-1996.  
PF 14-NOV-1994; 279537.  
PR 14-NOV-1994; JP-279537.  
PA (SUMO) SUMITOMO CHEM CO LTD.  
DR WPI; 96-318961/32.  
DR N-PSDB; T33315.  
PT Antibody recognising human-originated cytochrome P4501A2 -  
PT specifically recognises the species of P450 cytochrome present in a  
PT sample  
PS Example 1; Page 11-13; 15pp; Japanese.  
CC This sequence represents the cytochrome P4501A2. The cDNA sequence was  
CC amplified using the primers given in T3311-12. The human derived  
CC cytochrome P4501A2 was used in the generation of an antibody which  
CC is specific for this type of cytochrome. These antibodies may be  
CC used in the rapid and accurate determination of the exact cytochrome  
CC species present in a sample.  
SQ Sequence 516 AA;

Query Match 72.7%; Score 56; DB 1; Length 516;  
Best Local Similarity 85.7%; Pred. No. 3.96e+01;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 239 FPIVRYL 245  
QY 3 FPIVRYL 9  
|||||

RESULT 8  
ID R60101 standard; Protein; 713 AA.  
AC R60101;  
DE 15-MAR-1995 (first entry)  
DE Canine zona pellucida C2P2.  
KW Canine; dog; zona pellucida; 2P; C2P2; contraceptive; vaccine;  
KW antigen.  
OS Canis familiaris.  
PN J06189766-A.  
PD 12-JUL-1994.  
PF 25-DEC-1992; 359285.  
PR 25-DEC-1992; JP-359265.  
PA (TOFU) TONEN CORP.  
DR WPI; 94-259553/32.  
DR N-PSDB; Q70072.  
PT New DNA sequence encoding canine zona pellucida C2P2 - useful for  
PT the prodn. of a canine contraceptive vaccine antigen  
PS Claim 1; Page 8-10; 10pp; Japanese.  
CC The C2P2 DNA (Q70072) was prep'd. by the cloning of C2P2(75-520) -  
CC Q81700 using the primers given in Q70073-74, C2P2(1-65) - Q81804  
CC using the primers given in Q70082-83, C2P2(42-103) - Q81803 using the  
CC primers given in Q70079-81 and C2P2(487-713) - Q81957 using the  
CC primers given in Q70075-78.

DE Brushtail possum zona pellucida protein-2 (2P-2) protein.  
KW Zona pellucida protein; 2P-2; vaccine; female marsupial;  
KW contraceptive; conception; Brushtail possum; koala; kangaroo;  
KW wallaroo; wallaby; Pademelon.  
OS Trichosurus vulpecula.  
PN AU9878554-A.  
PD 11-FEB-1999.  
PF 29-JUL-1998; 078554.  
PR 12-FEB-1998; AU-001800.  
PR 31-JUL-1997; AU-008354.  
PA (MARS-) MARSUPIAL CRC LTD.  
PI Bradley M, Duckworth J, Mate K, McCartney C;  
DR WPI; 99-229776/20.  
DR N-PSDB; X34865.  
PT New marsupial zona pellucida (2P2 and 2P3) polypeptides for use in  
PT contraceptive vaccines  
PS Claim 20; Page 24-26; 43pp; English.  
CC The 2P-2 polypeptides or polynucleotides encoding them are  
CC administered as vaccines to female marsupials to raise an immune  
CC response against 2P-2 proteins and prevent conception. The population  
CC of koalas is growing which can cause death of food trees, and the  
CC Brushtail possum is New Zealand's number one vertebrate pest and can  
CC adversely affect the environment, animal health and the economy. Use  
CC of the new polypeptides as contraceptives can help control the  
CC population numbers of these and the Eastern grey and Western grey  
CC kangaroos, the Red kangaroo, the common wallaroo, Bennett's (or red  
CC necked) wallaby, the Tamar wallaby, the Whiptail wallaby, the Swamp  
CC wallaby, the Agile Wallaby and the Pademelon.  
SQ Sequence 712 AA;

Query Match 75.3%; Score 58; DB 1; Length 712;  
Best Local Similarity 85.7%; Pred. No. 2.41e+01;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

DB 510 YPIVRYL 516  
QY 3 FPIVRYL 9

RESULT 3  
ID R89370 standard; peptide; 9 AA.  
AC R89370;  
DE 18-SEP-1996 (first entry)  
DE Cw6 consensus peptide derived immunogenic peptide #2.  
KW Immunogenic peptide; supermotif; HLA molecule; CTL response;  
KW therapeutic; diagnostic; cancer; viral infection; hepatitis B;  
KW hepatitis C.  
OS Synthetic.  
PN W09603140-A1.  
PD 08-FEB-1996.  
PF 21-JUL-1995; U09234.  
PR 21-JUL-1994; US-278634.  
PR 23-NOV-1994; US-344824.  
PR 30-MAY-1995; US-452843.  
PA (CYTE-) CYTEL CORP.  
PI Sette A, Sidney J;  
DR WPI; 96-116784/12.  
PT Compn. comprising immunogenic peptide with supermotif allowing more  
PT than one HLA mol. to bind - used to induce CTL response in patient  
PT and for in vivo and ex vivo therapeutic and diagnostic applications  
PS Claim 26; Page 32pp; English.  
CC The sequences given in R89362-82 are immunogenic peptides which were  
CC use in the composition of the invention. The composition comprises  
CC an immunogenic peptide of 9-10 residues with a supermotif which  
CC allows binding of more than one HLA molecule. It pref. comprises  
CC two conserved residues, a first at the 2nd position from the N-  
CC terminal is Pro, and a 2nd at the C-terminal is Met. These peptides  
CC are used to induce a CTL response in a patient. They are also  
CC useful in compositions for in vivo and ex vivo therapeutic and  
CC diagnostic applications, e.g. the treatment of cancer and viral  
CC infections, e.g. hepatitis B and C.  
SQ Sequence 9 AA;

Query Match 74.0%; Score 57; DB 1; Length 9;  
Best Local Similarity 77.8%; Pred. No. 3.09e+01;  
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

DB 1 IPIPIVRSYL 9  
QY 1 IPIPIVRYL 9

## RESULT 4

ID W38927 standard; peptide; 15 AA.  
AC W38927;  
DE 27-MAR-1998 (first entry)  
DE Peptide resembling an SH3 domain binding peptide SEQ ID NO:324.  
KW Cortactin; SH3 domain; binding peptide; Src homology region 3;  
KW tyrosine kinase; immune response; lymphokine; interleukin 1; Nck;  
KW Abl; PLCgamma; p53bp2; Crk; Yes; Grb2.  
OS Synthetic.  
PN W09730074-A1.  
PD 21-AUG-1997.  
PF 14-FEB-1997; U02298.  
PR 16-FEB-1996; US-602999.  
PA (CYTO-) CYTOGEN CORP.  
PA (UYNC-) UNIV NORTH CAROLINA.  
PI Der CJ, Fowles DM, Kay BK, Quilliam LA, Rider JE,  
PI Sparks AB, Thorn JM;  
DR WPI; 97-424972/39.  
PT Src homology region 3 binding peptide - used to activate Src  
PT tyrosine kinase(s) and to stimulate immune response by increasing  
PT production of certain lymphokine(s), e.g. interleukin-1  
PS Claim 22; Page 90; 131pp; English.  
CC The present sequence represents a peptide which resembles a Src homology  
CC region 3 (SH3) binding peptide. SH3 binding peptides are selected from:  
CC (a) peptides which bind the SH3 domain of Cortactin; (b) peptides which  
CC bind the middle SH3 domain of Nck; (c) peptides which bind the SH3  
CC domain of Abl; (d) peptides which bind the SH3 domain of Src; (e)  
CC peptides which bind the SH3 domain of PLC gamma; (f) peptides which bind  
CC the SH3 domain of p53bp2; (g) peptides which bind the amino-terminal SH3  
CC domain of Crk; (h) peptides which bind the SH3 domain of Yes; and (i)  
CC peptides which bind the amino-terminal SH3 domain of Grb2. The purified  
CC binding peptides can be used in the method to identify inhibitors of  
CC their binding to their respective SH3 domains, which could be used to  
CC modulate the pharmacological activity of proteins or polypeptide  
CC containing the SH3 domain. The peptides can also be used to activate  
CC Src or Src-related protein tyrosine kinases, to stimulate the immune  
CC response by increasing the production of certain lymphokines, e.g.  
CC tumour necrosis factor-alpha and interleukin-1, or to deliver a  
CC conjugated molecule to certain cellular compartments containing Src or  
CC Src related proteins.  
SQ Sequence 15 AA;

Query Match 72.7%; Score 56; DB 1; Length 15;  
Best Local Similarity 85.7%; Pred. No. 3.96e+01;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

DB 1 FPIVRYL 7  
QY 3 FPIVRYL 9

## RESULT 5

ID R93167 standard; Protein; 516 AA.  
AC R93167;  
DE 11-OCT-1996 (first entry)  
DE Human cytochrome P450 molecular species 1A2 protein.  
KW Human cytochrome P450; amplified; PCR; polymerase chain reaction; primer;  
KW liver; yeast; expression vector; NADPH-P450 reductase; ADH gene promoter;  
KW evaluation; safety; fusion protein; metabolite; detoxification;  
KW carcinogenic.  
KW Homo sapiens.  
PN J08056655-A.  
PD 05-MAR-1996.

\*\*\*\*\*  
WIREH  
\*\*\*\*\*  
(TW)

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Fri Apr 14 23:54:28 2000; Maspar time 6.28 seconds  
33.921 Million cell updates/sec  
Tabular output not generated.

Title: >US-08-452-843-10  
Description: (1-9) from US08452843.pep  
Perfect Score: 77  
Sequence: 1 IPPFIVRYL 9

Scoring table: PAM 150  
Gap 15

Searched: 188963 seqs, 23686106 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: a-geneseq36  
1:geneseqp

Statistics: Mean 17.954; Variance 54.576; scale 0.329

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	77	100.0	9	1 R89371	Cw6 consensus peptide	1.74e+01
2	58	75.3	712	1 Y01772	Brush tail possum zona	2.41e+01
3	57	74.0	9	1 R89370	Cw6 consensus peptide	3.09e+01
4	56	72.7	15	1 W38927	Peptide resembling an	3.96e+01
5	56	72.7	516	1 R93167	Human cytochrome P450	3.96e+01
6	56	72.7	516	1 R72360	Human cytochrome P450	3.96e+01
7	56	72.7	516	1 W00183	Cytochrome P4501A2	3.96e+01
8	56	72.7	713	1 R60101	Canine zona pellucida	3.96e+01
9	56	72.7	715	1 W81808	Canine zona pellucida	3.96e+01
10	56	72.7	715	1 R55198	Canine zona pellucida	3.96e+01
11	56	72.7	716	1 R55200	Feline zona pellucida	3.96e+01
12	56	72.7	716	1 W81810	Feline zona pellucida	3.96e+01
13	56	72.7	716	1 R60532	Feline zona pellucida	3.96e+01
14	53	68.8	9	1 R89369	Cw6 consensus peptide	8.25e+01
15	53	68.8	713	1 R08998	Mouse ZP2 protein exhi	8.25e+01
16	51	66.2	271	1 W19219	Human growth hormone s	1.34e+02
17	51	66.2	271	1 W19612	Human growth hormone s	1.34e+02
18	51	66.2	341	1 R87029	Varicella zoster virus	1.34e+02
19	51	66.2	341	1 R74389	VZV thymidine-kinase	1.34e+02
20	51	66.2	341	1 R74390	VZV thymidine-kinase	1.34e+02
21	51	66.2	341	1 R74391	VZV thymidine-kinase	1.34e+02
22	51	66.2	353	1 W19215	Swine growth hormone s	1.34e+02
23	51	66.2	353	1 W19608	Pig growth hormone sec	1.34e+02

24	51	66.2	361	1 W19217	Human growth hormone s	1.34e+02
25	51	66.2	362	1 W19610	Human growth hormone s	1.34e+02
26	51	66.2	364	1 W19220	Rat growth hormone sec	1.34e+02
27	51	66.2	364	1 W19613	Rat growth hormone sec	1.34e+02
28	51	66.2	380	1 R37595	Sequence of microsomal	1.34e+02
29	51	66.2	463	1 R41877	Rat glucagon-like pept	1.34e+02
30	51	66.2	713	1 W81804	Porcine ZPA protein	1.34e+02
31	51	66.2	713	1 R55194	Porcine zona pellucida	1.34e+02
32	50	64.9	300	1 R76772	FimH protein derived f	1.70e+02
33	50	64.9	398	1 W1368	Death associated prote	1.70e+02
34	50	64.9	398	1 R74206	Human death associated	1.70e+02
35	50	64.9	575	1 Y00157	Enterococcus faecalis	1.70e+02
36	50	64.9	601	1 Y00156	Enterococcus faecalis	1.70e+02
37	49	63.6	416	1 W54098	Homo sapiens B15 sequ	2.15e+02
38	49	63.6	463	1 R70006	Human glucagon-like 1	2.15e+02
39	49	63.6	644	1 W82318	Human 7-transmembrane	2.15e+02
40	49	63.6	951	1 W93965	Human AOMF05 protein	2.15e+02
41	49	63.6	951	1 W93906	Human AOMF05 protein	2.15e+02
42	48	62.3	125	1 R38224	Sequence of polypeptid	2.72e+02
43	48	62.3	210	1 R13499	P.denitrificans COB H	2.72e+02
44	48	62.3	742	1 R74094	Human zona pellucida-2	2.72e+02
45	48	62.3	745	1 R55206	Human zona pellucida 2	2.72e+02

ALIGNMENTS

RESULT 1  
ID R89371 standard; peptide; 9 AA.  
AC R89371; 1996 (first entry)  
DT 18-SEP-1996  
DE Cw6 consensus peptide derived immunogenic peptide #3.  
KW Immunogenic peptide; supermotif; HLA molecule; CTL response;  
KW therapeutic; diagnostic; cancer; viral infection; hepatitis B;  
KW hepatitis C.  
OS Synthetic.  
FN WO9603140-A1.  
PD 08-FEB-1996.  
PF 21-JUL-1995; U09234.  
PR 21-JUL-1994; US-278634.  
PR 23-NOV-1994; US-344824.  
PR 30-MAY-1995; US-452843.  
FA (CYTE-) CYTEL CORP.  
PI Sette A, Sidney J.  
DR WPI; 96-116784/12.  
PT Compsn. comprising immunogenic peptide with supermotif allowing more than one HLA mol. to bind - used to induce CTL response in patient and for in vivo and ex vivo therapeutic and diagnostic applications  
PS Claim 2; Page 26; 32pp; English.  
CC The sequences given in R89362-82 are immunogenic peptides which were use in the composition of the invention. The composition comprises an immunogenic peptide of 9-10 residues with a supermotif which allows binding of more than one HLA molecule. It pref. comprises two conserved residues, a first at the 2nd position from the N-terminal is Pro, and a 2nd at the C-terminal is Met. These peptides are used to induce a CTL response in a patient. They are also useful in compositions for in vivo and ex vivo therapeutic and diagnostic applications, e.g. the treatment of cancer and viral infections, e.g. hepatitis B and C.  
SQ Sequence 9 AA;  
Query Match 100.0%; Score 77; DB 1; Length 9;  
Best Local Similarity 100.0%; Pred. No. 1.74e+01;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Db 1 IPPFIVRYL 9  
Qy 1 IPPFIVRYL 9  
RESULT 2  
ID Y01772 standard; Protein; 712 AA.  
AC Y01772;  
DT 28-JUN-1999 (first entry)

THIS PAGE BLANK (USPTO)

RA WASHU;  
 RT "The A. thaliana Genome Sequencing Project."  
 RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. COLOMBIA;  
 RA RYAN E., EDWARDS J., PAPE K.;  
 RT "The sequence of A. thaliana F6N15."  
 RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. COLOMBIA;  
 RA WATERSTON R.;  
 RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF069299; AAC19309.1;  
 SQ SEQUENCE 1260 AA; 139284 MW; 29D67BA5 CRC32;

Query Match 71.4%; Score 50; DB 10; Length 1260;  
 Best Local Similarity 55.6%; Pred. No. 2.22e+01;  
 Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 784 LPYPIRLPI 792  
 QY 1 IPYPIVRS 9

RESULT 13  
 ID Q14991 PRELIMINARY; PRT; 57 AA.  
 AC Q14991;  
 DT 01-NOV-1996 (TRENBLrel. 01, Created)  
 DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)  
 DT 01-NOV-1998 (TRENBLrel. 08, Last annotation update)  
 DE PROTEIN OF UNKNOWN FUNCTION.  
 OS Homo sapiens (Human)  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=OVARY;  
 RX MEDLINE; 91025550.  
 RA RAPP G., FREUDENSTEIN J., KLAUDINY J., MUCHA J., WEMPE F., ZIMMER M.,  
 RA SCHEIT K.H.;  
 RT "Characterization of three abundant mRNAs from human ovarian granulosa cells."  
 RL DNA Cell Biol. 9:479-485(1990).  
 DR EMBL: M38188; AAA63233.1;  
 SQ SEQUENCE 57 AA; 6834 MW; 9F00B7D3 CRC32;

Query Match 70.0%; Score 49; DB 4; Length 57;  
 Best Local Similarity 75.0%; Pred. No. 3.49e+01;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 34 PYPPIGRSM 41  
 QY 2 YPIVRS 9

RESULT 14  
 ID P74029 PRELIMINARY; PRT; 219 AA.  
 AC P74029;  
 DT 01-FEB-1997 (TRENBLrel. 02, Created)  
 DT 01-FEB-1997 (TRENBLrel. 02, Last sequence update)  
 DT 01-JAN-1999 (TRENBLrel. 09, Last annotation update)  
 DE HYPOTHETICAL 23.5 KD PROTEIN.  
 GN YCF39.  
 OS Synechocystis sp. (strain PCC 6803).  
 OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=PCC6803;  
 RA TABATA S.;  
 RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.  
 RN [2]

RP SEQUENCE FROM N.A.  
 RC STRAIN=PCC6803;  
 RX MEDLINE; 97061201.  
 RA KANEKO T., SATO S., KOTANI H., TANAKA A., ASAMIZU E., NAKAMURA Y.,  
 RA MIYAJIMA N., HIROSAWA M., SUGIURA M., SASAMOTO S., KIMURA T.,  
 RA HOSOUCHI T., MATSUNO A., MURAKI A., NAKAZAKI N., NAROO K., OKUMURA S.,  
 RA SHIMPO S., TAKEUCHI C., WADA T., WATANABE A., YAMADA M., YASUDA M.,  
 RA TABATA S.;  
 RT "Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis sp. strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions."  
 RL DNA Res. 3:109-136(1996).  
 DR EMBL: D90911; BAA18102.1;  
 KW Hypothetical protein.  
 SQ SEQUENCE 219 AA; 23534 MW; 37A9EA9C CRC32;

Query Match 70.0%; Score 49; DB 2; Length 219;  
 Best Local Similarity 62.5%; Pred. No. 3.49e+01;  
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 143 VPYTIVRP 150  
 QY 1 IPYPIVRS 8

RESULT 15  
 ID Q9WGH9 PRELIMINARY; PRT; 301 AA.  
 AC Q9WGH9;  
 DT 01-NOV-1999 (TRENBLrel. 12, Created)  
 DT 01-NOV-1999 (TRENBLrel. 12, Last sequence update)  
 DT 01-NOV-1999 (TRENBLrel. 12, Last annotation update)  
 DE GAG PROTEIN (FRAGMENT).  
 GN GAG.  
 OS Human immunodeficiency virus type 1.  
 OC Viruses; Retroviridae; Retroviridae; Lentivirus.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=08102-2-GAGA;  
 RX MEDLINE; 99214336.  
 RA WILSON C.C., BROWN R.C., KORBER B.T., WILKES B.M., RUHL D.J.,  
 RA SARAKOTO D., KUNSTMAN K., LUZURIAGA K., HANSON I.C., WIDMAYER S.M.,  
 RA WIGNIA A., CLAPP S., AMMANN A.J., KOUP R.A., WOLINSKY S.M.,  
 RA WALKER B.D.;  
 RT "Frequent detection of escape from cytotoxic T-lymphocyte recognition in perinatal human immunodeficiency virus (HIV) type 1 transmission: the ariel project for the prevention of transmission of HIV from mother to infant."  
 RL J. Virol. 73:3975-3985(1999).  
 DR EMBL: AF121482; AAD28854.1;  
 FT NON\_TER 1  
 FT NON\_TER 301 301  
 SQ SEQUENCE 301 AA; 33737 MW; D5F71845 CRC32;

Query Match 70.0%; Score 49; DB 14; Length 301;  
 Best Local Similarity 85.7%; Pred. No. 3.49e+01;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 132 YPIVRSI 138  
 QY 3 YPIVRS 9

Search completed: Fri Apr 14 23:52:41 2000  
 Job time : 103 secs.



!

Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 71 IPYIAKSL 79  
|||||:  
Qy 1 IPYIVRSL 9

RESULT 6  
ID O33331 PRELIMINARY; PRT; 410 AA.  
AC O33331  
DT 01-JAN-1998 (TReMBLrel. 05, Created)  
DT 01-JAN-1998 (TReMBLrel. 05, Last sequence update)  
DT 01-NOV-1999 (TReMBLrel. 12, Last annotation update)  
DE DEHYDROGENASE.  
GN MTV002.54C.  
OS Mycobacterium tuberculosis.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
RN [1]  
RP STRAIN-H37RV;  
RC STRAIN-H37RV;  
RA MURPHY L., HARRIS D.;  
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-H37RV;  
RA PARKHILL J., BARRELL B.G., RAJANDREAM M.A.;  
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-H37RV;  
RA PHILIP W.J., POULET S., EIGLMEIER K., PASCOPELLA L.,  
RA BALASUBRAMANIAN V., HEYM B., BERGH S., BLOOM B.R., JACOBS W.R. JR.,  
RA COLE S.T.;  
RT "An integrated map of the genome of the tubercle bacillus,  
RT Mycobacterium tuberculosis H37RV, and comparison with Mycobacterium  
RT leprae.";  
RL Proc. Natl. Acad. Sci. U.S.A. 93:3132-3137(1996).  
DR EMBL; AL008967; CAAL5584.1; -  
DR HSP; Q06319; IBCU.  
DR PFAM; PF00441; Acyl-CoA\_dh; 1.  
SQ SEQUENCE 410 AA; 44743 MW; 98A843CB CRC32;

Query Match 72.9%; Score 51; DB 2; Length 410;  
Best Local Similarity 75.0%; Pred. No. 1.41e+01;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 37 PYPIARKL 44  
|||||:  
Qy 2 PYPIVRSL 9

RESULT 7  
ID O34726 PRELIMINARY; PRT; 478 AA.  
AC O34726;  
DT 01-JAN-1998 (TReMBLrel. 05, Created)  
DT 01-JAN-1998 (TReMBLrel. 05, Last sequence update)  
DT 01-NOV-1998 (TReMBLrel. 08, Last annotation update)  
DE YFUS PROTEIN.  
GN YFUS.  
OS Bacillus subtilis.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
OC Bacillus/Staphylococcus group; Bacillus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-168;  
RX MEDLINE; 98044033.  
RA KUNST F., OGASAWARA N., MOSZER I., ALBERTINI A.M., ALLONI G.,  
RA AZEVEDO V., BERRERO M.G., BESSIERES P., BOLOTIN A., BORCHERT S.,  
RA BOKRIS R., BOURSIER L., BRANS A., BRAUN N., BRIGHELL S.C., BRON S.,  
RA BROUILLET S., BRUSCHI C.V., CALDWELL B., CAPUANO V., CARTER N.M.,  
RA CHOI S.K., CODANI J.J., CONNERTON I.F., CUMMINGS N.J., DANIEL R.A.,

RA DENIZOT F., DEVINE K.M., DUSTERHOFT A., EHRLICH S.D., EMMERSON P.T.,  
RA ENTIAN K.D., ERRINGTON J., FABRET C., FERRARI E., FOULGER D.,  
RA FRITZ C., FUJITA M., FUJITA Y., FUMA S., GALIZZI A., GALLERON N.,  
RA GHIM S.Y., GLASER P., GORFEAU A., GOLIGHTLY E.J., GRANDI G.,  
RA GUISEPPI G., GUY B.J., HAGA K., HATICH J., HARWOOD C.R., HENAUT A.,  
RA HILBERT H., HOLSAPPEL S., HOSONO S., HULLO M.F., ITAYA M., JONES L.,  
RA JORIS B., KARAMATA D., KASAHARA Y., KLAERR-BLANCHARD M., KLEIN C.,  
RA KOBAYASHI Y., KOETTER P., KONINGSTEIN G., KROGH S., KUMANO M.,  
RA KURITA K., LAPIDUS A., LARDINOIS S., LAUBER J., LAZAREVIC V.,  
RA LEE S.M., LEVINE A., LIU H., MASUDA S., MAUEL C., MEDIGUE C.,  
RA MEDINA N., MELLADO R.P., MIZUNO M., MOESTL D., NAKAI S., NOBACK M.,  
RA NOONE D., O'REILLY M., OGAWA K., OGIWARA A., OUDEGA B., PARK S.H.,  
RA PARRO V., POHL T.M., PORTELELLA D., PORMOLLIK S., PRESCOTT A.M.,  
RA PRESECAN E., PUJIC P., PURNELLE B., RAPOPORT G., REY M., REYNOLDS S.,  
RA RIEGER M., RIVOLTA C., ROCHA E., ROCHE B., ROSE M., SADAIE Y.,  
RA SATO T., SCANLAN E., SCHLEICH S., SCHROETER R., SCORFONE F.,  
RA SEKIGUCHI J., SEKONSKA A., SERO S.J., SERROR P., SHIN B.S., SOLDI B.,  
RA SOROKIN A., TACCONI E., TAKAGI T., TAKAHASHI H., TAKEKAWA K.,  
RA TAKEUCHI M., TAKAKOSHI A., TANAKA T., TERPSTRA P., TOGNONI A.,  
RA TOSATO V., UCHIYAMA S., VANDENBOL M., VANNIER F., VASSAROTTI A.,  
RA VIARI A., WAMBUIT R., WEDLER E., WEDLER H., WEITZENEGER T.,  
RA WINTERS P., WIPAT A., YAMAMOTO H., YAMANE K., YASUMOTO K., YATA K.,  
RA YOSHIDA K., YOSHIKAWA H.F., ZUMSTEIN E., YOSHIKAWA H., DANCHIN A.;  
RT "The complete genome sequence of the gram-positive bacterium Bacillus  
RT subtilis.";  
RL Nature 390:249-256(1997).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-168;  
RA KUNST F., OGASAWARA N., YOSHIKAWA H., DANCHIN A.;  
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-AC327;  
RX MEDLINE; 97417488.  
RA YAMAMOTO H., UCHIYAMA S., NUGROHO F.A., SEKIGUCHI J.;  
RT "Cloning and sequencing of a 35.7 kb in the 70 degree-73 degree region  
RT of the Bacillus subtilis genome reveal genes for a new two-component  
RT system, three spore germination proteins, an iron uptake system and a  
RT general stress response protein.";  
RL Gene 194:191-199(1997).  
DR EMBL; 295108; CAB12586.1; -  
DR EMBL; D86417; BAA2312.1; -  
DR PFAM; PF00939; Na\_sulph\_symp; 1.  
SQ SEQUENCE 478 AA; 51431 MW; 90C9082D CRC32;

Query Match 72.9%; Score 51; DB 2; Length 478;  
Best Local Similarity 66.7%; Pred. No. 1.41e+01;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 155 IIPPIIRSL 163  
|:|:|:  
Qy 1 IIPYIVRSL 9

RESULT 8  
ID O66834 PRELIMINARY; PRT; 520 AA.  
AC O66834;  
DT 01-AUG-1998 (TReMBLrel. 07, Created)  
DT 01-AUG-1998 (TReMBLrel. 07, Last sequence update)  
DT 01-NOV-1998 (TReMBLrel. 08, Last annotation update)  
DE RECOMBINATION PROTEIN RECN.  
GN RECN.  
OS Aquifex aeolicus.  
OC Bacteria; Aquificales; Aquificaceae; Aquifex.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-VF5;  
RX MEDLINE; 98196666.  
RA DECKERT G., WARREN P.V., GAASTERLAND T., YOUNG W.G., LENOX A.L.,  
RA GRAHAM D.E., OVERBEEK R., SNEAD M.A., KELLER M., AUSTAY M., HUBER R.,  
RA FELDMAN R.A., SHORT J.M., OLSON G.J., SWANSON R.V.;  
RT "The complete genome of the hyperthermophilic bacterium Aquifex

```

ID Q84619 PRELIMINARY; PRT; 112 AA.
AC Q84619;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE GENOME, PARTIAL SEQUENCE.
GN A303L.
OS Paramyxium bursaria chlorella virus 1 (PBCV-1).
OC Viruses; dsDNA viruses, no RNA stage; Phycodnaviridae; Phycodnavirus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 95133167.
RA LU Z., LI Y., ZHANG Y., KUTISH G.F., ROCK D.L., VAN ETTEN J.L.;
RT "Analysis of 45 kb of DNA located at the left end of the chlorella
RT virus PBCV-1 genome.";
RL Virology 206:339-352(1995).
DR EMBL; U42580; AAC96671.1; -.
SQ SEQUENCE 112 AA; 13416 MW; 5C07006C CRC32;

Query Match 72.9%; Score 51; DB 14; Length 112;
Best Local Similarity 44.4%; Pred. No. 1.41e+01;
Matches 4; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

Db 20 VPFSIIRNL 28
QY 1 IPYIVRSL 9

RESULT 3 PRELIMINARY; PRT; 300 AA.
AC Q9WX26;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE HYPOTHETICAL 33.1 KD PROTEIN.
GN SCE68.05C.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA MURPHY L., HARRIS D.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA JAMES K.D., PARKHILL J., BARRELL B.G., RAJANDREAM M.A.;
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RX MEDLINE; 97000351.
RA REDENBACH M., KIESER H.M., DENAPATE D., EICHNER A., CULLUM J.,
RA KINASHI H., HOPWOOD D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
DR EMBL; AL079345; CAB45341.1; -.
KW Hypothetical protein.
SQ SEQUENCE 300 AA; 33068 MW; F98702D5 CRC32;

Query Match 72.9%; Score 51; DB 2; Length 300;
Best Local Similarity 75.0%; Pred. No. 1.41e+01;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 176 APYIVREL 183
QY 2 PYIVRSL 9

```

```

RESULT 4 PRELIMINARY; PRT; 305 AA.
ID Q9YBQ5;
AC Q9YBQ5;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE 305AA LONG HYPOTHETICAL PROTEIN.
GN APE1544.
OS Aeropyrum pernix.
OC Archaea; Crenarchaeota; Aeropyrum.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 99310339.
RA KAWARABAYASHI Y., HINO Y., HORIKAWA H., YAMAZAKI S., HAIKAWA Y.,
RA JIN-NO K., TAKAHASHI M., SEKINE M., BABA S., ANKAI A., KOSUGI H.,
RA HOSOYAMA A., FUKUI S., NAGAI Y., NISHIJIMA K., NAKAZAWA H.,
RA TAKAMIYA M., MASUDA S., FUNAHASHI T., TANAKA T., KUDOH Y.,
RA YAMAZAKI J., KUSHIDA N., OGUCHI A., AOKI K., KUBOTA K., NAKAMURA Y.,
RA NOMURA N., SAKO Y., KIKUCHI H.;
RT "Complete genome sequence of an aerobic hyper-thermophilic
RT crenarchaeon, Aeropyrum pernix K1.";
RL DNA Res. 6:83-101(1999).
DR EMBL; AP000061; BAA80543.1; -.
SQ SEQUENCE 305 AA; 32739 MW; BB8A06B1 CRC32;

Query Match 72.9%; Score 51; DB 1; Length 305;
Best Local Similarity 71.4%; Pred. No. 1.41e+01;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 91 YPVVRAL 97
QY 3 YPIVRSL 9

RESULT 5 PRELIMINARY; PRT; 398 AA.
ID Q07717;
AC Q07717;
DT 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE ACEAB.
GN ACEAB.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-H37RV;
RA OLIVER K., HARRIS D.;
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-H37RV;
RA BARRELL B.G., RAJANDREAM M.A., PARKHILL J.;
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-H37RV;
RX MEDLINE; 96181548.
RA PHILIPP W.J., POULET S., EIGLMEIER K., PASCOPELLA L., JACOBS W.R. JR.,
RA BALASUBRAMANIAN V., HEYM B., BERGH S., BLOOM B.R.,
RA COLE S.T.;
RT "An integrated map of the genome of the tubercle bacillus,
RT Mycobacterium tuberculosis H37Rv, and comparison with Mycobacterium
RT leprae.";
RL Proc. Natl. Acad. Sci. U.S.A. 93:3132-3137(1996).
DR EMBL; Z97193; CAB10026.1; -.
DR PFAM; PF00463; ICL; 1.
SQ SEQUENCE 398 AA; 44581 MW; 95F2E718 CRC32;

Query Match 72.9%; Score 51; DB 2; Length 398;
Best Local Similarity 66.7%; Pred. No. 1.41e+01;

```

\*\*\*\*\*  
MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Fri Apr 14 23:50:58 2000; MasPar time 13.13 Seconds  
Tabular output not generated.  
47.542 Million cell updates/sec.  
Title: >US-08-452-843-9  
Description: (1-9) from US08452843.pep  
Perfect Score: 70  
Sequence: 1 IPYPIVRSLS 9  
Scoring table: PAM 150  
Gap 15  
Searched: 225878 seqs, 69334122 residues  
Post-processing: Minimum Match 0%  
Listing first 45 summaries  
Database: sptrembl12  
1:sp-archaea 2:sp-bacteria 3:sp-fungi 4:sp-human  
5:sp-invertebrate 6:sp-mammal 7:sp-mhc 8:sp-organellar  
9:sp-phage 10:sp-plant 11:sp-rodent 12:sp-unclassified  
13:sp-vertebrate 14:sp-virus  
Statistics: Mean 24.245; Variance 29.817; scale 0.813  
Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.  
SUMMARIES  
Result No. Score Query Match Length DB ID Description Pred. No.  
1 52 74.3 344 10 Q92PF5 T3H13.10 PROTEIN. 8.85e+00  
2 51 72.9 112 14 Q84519 GENOME, PARTIAL SEQUEN 1.41e+01  
3 51 72.9 300 2 Q9W26 HYPOTHETICAL 33.1 KD P 1.41e+01  
4 51 72.9 305 1 Q9YB05 305AA LONG HYPOTHETICA 1.41e+01  
5 51 72.9 398 2 Q07717 ACEAB. 1.41e+01  
6 51 72.9 410 2 Q33331 DEHYDROGENASE. 1.41e+01  
7 51 72.9 478 2 Q34726 YFLS PROTEIN. 1.41e+01  
8 51 72.9 520 2 Q66834 RECOMBINATION PROTEIN 1.41e+01  
9 50 71.4 337 10 Q80934 F13M22.16 PROTEIN. 2.22e+01  
10 50 71.4 486 9 Q21944 GLUCOSYL TRANSFERASE II 2.22e+01  
11 50 71.4 537 2 Q45977 RNASE E (FRAGMENT). 2.22e+01  
12 50 71.4 1260 10 Q81307 F6N15.10 PROTEIN. 2.22e+01  
13 49 70.0 57 4 Q14991 PROTEIN OF UNKNOWN FUN 3.49e+01  
14 49 70.0 219 2 P74029 HYPOTHETICAL 23.5 KD P 3.49e+01  
15 49 70.0 301 14 Q9WGH9 GAG PROTEIN (FRAGMENT) 3.49e+01  
16 49 70.0 309 4 Q95010 WUGSC:H\_DJ0728D04.1 PR 3.49e+01  
17 49 70.0 460 2 Q925P6 FUMARATE HYDRATASE. 3.49e+01  
18 49 70.0 502 2 Q9X231 LYSYL-TRNA SYNTHETASE. 3.49e+01  
19 49 70.0 565 2 Q926P5 SULFATE TRANSPORTER. 3.49e+01  
20 49 70.0 567 2 Q84864 SULFATE TRANSPORTER. 3.49e+01

\*\*\*\*\*  
Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd  
\*\*\*\*\*  
(TM)

\*\*\*\*\*  
21 49 70.0 667 5 045995 2K1053.4 PROTEIN. 3.49e+01  
22 48 68.6 293 14 Q9WGN0 GAG PROTEIN (FRAGMENT) 5.43e+01  
23 48 68.6 371 4 Q43190 PURINERGIC P2Y11 RECP 5.43e+01  
24 48 68.6 432 2 Q34978 YTIIP. 5.43e+01  
25 48 68.6 445 1 Q9YDV0 445AA LONG HYPOTHETICA 5.43e+01  
26 48 68.6 451 2 P74054 HYPOTHETICAL 50.4 KD P 5.43e+01  
27 48 68.6 530 4 Q9YSP3 RETINOIC ACID-INDUCED 5.43e+01  
28 48 68.6 535 5 Q19862 F28C6.2 PROTEIN. 5.43e+01  
29 48 68.6 764 11 Q54860 CARBOXYPEPTIDASE X2. 5.43e+01  
30 48 68.6 1026 3 Q06315 CHROMOSOME XII COSMID 5.43e+01  
31 48 68.6 1058 10 Q9XIK4 T10024.10. 5.43e+01  
32 48 68.6 1339 10 Q82180 T4C15.22 PROTEIN. 5.43e+01  
33 48 68.6 1413 5 Q23596 ZK732.1 PROTEIN. 5.43e+01  
34 47 67.1 129 14 Q36882 MA-P17 (FRAGMENT). 8.40e+01  
35 47 67.1 132 14 Q36877 MA-P17 (FRAGMENT). 8.40e+01  
36 47 67.1 212 2 Q86043 GLUTATHIONE-S-TRANSFER 8.40e+01  
37 47 67.1 228 1 P95967 ORF C04028 8.40e+01  
38 47 67.1 352 14 Q9YMN5 CAPSID PROTEIN. 8.40e+01  
39 47 67.1 568 2 Q929G0 PREDICTED OMP. 8.40e+01  
40 47 67.1 588 10 P93341 PHOSPHOINOSITIDE-SPECI 8.40e+01  
41 47 67.1 588 10 Q49902 1-PHOSPHATIDYLINOSITOL 8.40e+01  
42 47 67.1 665 5 Q60966 PCACA. 8.40e+01  
43 47 67.1 712 6 Q77685 ZONA PELLUCIDA 2 PROTE 8.40e+01  
44 47 67.1 1139 4 Q15073 130K PROTEIN (PRB2/PI3 8.40e+01  
45 47 67.1 1139 4 Q16084 PI30. 8.40e+01

ALIGNMENTS

RESULT 1  
ID Q92PF5 PRELIMINARY; PRT; 344 AA.  
AC Q92PF5;  
DT 01-MAY-1999 (TREMBlrel. 10, Created)  
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)  
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)  
DE T3H13.10 PROTEIN.  
GN T3H13.10.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC euryhalophytes; Spermatophyta; Magnoliophyta; eudicotyledons;  
OC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;  
OC Arabidopsis.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV. COLUMBIA;  
RA WASHU;  
RT "The A. thaliana Genome Sequencing Project";  
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV. COLUMBIA;  
RA DRONE K., NGUYEN C.;  
RT "The sequence of A. thaliana T3H13.";  
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV. COLUMBIA;  
RA WATERSTON R.;  
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF128396; AAD17371.1; -;  
DR MENDEL; 40010; Arath;2050;40010.  
SQ SEQUENCE 344 AA; 37980 MW; C8D86D1D CRC32;

Query Match 74.3%; Score 52; DB 10; Length 344;  
Best Local Similarity 55.6%; Pred. No. 8.85e+00;  
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;  
Db 165 ISYPIVRSLS 173  
QY 1 IPYPIVRSLS 9  
RESULT 2

**THIS PAGE BLANK (USPTO)**

CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; U67468; AAB98088.1; -  
 DR TIGR; M30106; -  
 KW Hypothetical protein; Transmembrane.  
 FT TRANSMEM 19 39 POTENTIAL.  
 FT TRANSMEM 79 99 POTENTIAL.  
 FT TRANSMEM 141 161 POTENTIAL.  
 SQ SEQUENCE 238 AA; 26473 MW; 6443385A CRC32;

Query Match 70.0%; Score 49; DB 1; Length 238;  
 Best Local Similarity 75.0%; Pred. No. 8.36e+00;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 65 IPYPPVRA 72  
 Qy 1 IPYPIVRS 8

Search completed: Fri Apr 14 23:50:40 2000  
 Job time : 48 secs.

Best Local Similarity 75.0%; Pred. No. 5.12e+00;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 394 PYPKRAL 401  
QY 2 PYPVRSLS 9  
|||||:|

RESULT 13  
ID DYHC-FUSSO STANDARD; PRT; 4349 AA.  
AC P8716;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE DNEIN HEAVY CHAIN, CYTOSOLIC (DYHC).  
GN DCL1.  
OS Fusarium solani (subsp. pisi) (Nectria haematococca).  
OC Eukaryota; Fungi; Ascomycota; Euascomycetes; Pyrenomycetes;  
OC Hypocreales; Hypocreaceae; Nectria.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=T213;  
RA INOUE S., AIST J.R., TURGEON B.G., YODER O.C.;  
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: DYNEIN HAS ATPASE ACTIVITY. CYTOPLASMIC DYNEIN ACTS AS A  
CC MOTOR FOR THE INTRACELLULAR RETROGRADE MOTILITY OF VESICLES AND  
CC ORGANELLES ALONG MICROTUBULES.  
CC -!- SUBUNIT: CONSISTS OF AT LEAST TWO HEAVY CHAINS AND A NUMBER OF  
CC INTERMEDIATE AND LIGHT CHAINS.  
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.  
CC -!- SIMILARITY: BELONGS TO THE DYNEIN HEAVY CHAIN FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; U84215; AAC33176.1; -  
DR HSSP; P03069; 1ZIJ.  
KW Motor protein; Microtubules; Dynein; ATP-binding;  
KW Heptad repeat pattern.  
FT NP\_BIND 1946 1953 ATP (POTENTIAL).  
FT NP\_BIND 2239 2246 ATP (POTENTIAL).  
FT NP\_BIND 2604 2611 ATP (POTENTIAL).  
FT NP\_BIND 2946 2953 ATP (POTENTIAL).  
FT NP\_BIND 4349 AA; 493453 MW; 961A2CID CRC32;  
SQ SEQUENCE 4349 AA; 493453 MW; 961A2CID CRC32;

Query Match 71.4%; Score 50; DB 1; Length 4349;  
Best Local Similarity 75.0%; Pred. No. 5.12e+00;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 407 PYPKRAL 414  
QY 2 PYPVRSLS 9  
|||||:|

RESULT 14  
ID DYHC-NEUCR STANDARD; PRT; 4367 AA.  
AC P45443;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE DYNEIN HEAVY CHAIN, CYTOSOLIC (DYHC).  
GN RO-1.  
OS Neurospora crassa.  
OC Eukaryota; Fungi; Ascomycota; Euascomycetes; Pyrenomycetes;  
OC Sordariales; Sordariaceae; Neurospora.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=74-OR23-1A;

MEDLINE; 95014704.  
RX PLAMANN M., MINKE P.F., TINSLEY J.H., BRUNO K.S.;  
RA "Cytoplasmic dynein and actin-related protein Arp1 are required for  
RT normal nuclear distribution in filamentous fungi";  
RL J. Cell Biol. 127:139-149(1994).  
CC -!- FUNCTION: DYNEIN HAS ATPASE ACTIVITY. CYTOPLASMIC DYNEIN ACTS AS A  
CC MOTOR FOR THE INTRACELLULAR RETROGRADE MOTILITY OF VESICLES AND  
CC ORGANELLES ALONG MICROTUBULES. REQUIRED TO MAINTAIN UNIFORM  
CC NUCLEAR DISTRIBUTION IN HYPHAE.  
CC -!- SUBUNIT: CONSISTS OF AT LEAST TWO HEAVY CHAINS AND A NUMBER OF  
CC INTERMEDIATE AND LIGHT CHAINS.  
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.  
CC -!- SIMILARITY: BELONGS TO THE DYNEIN HEAVY CHAIN FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; L31504; AAA64908.1;  
DR Motor protein; Microtubules; Dynein; ATP-binding;  
KW Heptad repeat pattern.  
FT NP\_BIND 1943 1950 ATP (POTENTIAL).  
FT NP\_BIND 2240 2247 ATP (POTENTIAL).  
FT NP\_BIND 2605 2612 ATP (POTENTIAL).  
FT NP\_BIND 2947 2954 ATP (POTENTIAL).  
FT NP\_BIND 4367 AA; 495568 MW; B81B5E92 CRC32;  
SQ SEQUENCE 4367 AA; 495568 MW; B81B5E92 CRC32;

Query Match 71.4%; Score 50; DB 1; Length 4367;  
Best Local Similarity 75.0%; Pred. No. 5.12e+00;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 404 PYPKRAL 411  
QY 2 PYPVRSLS 9  
|||||:|

RESULT 15  
ID Y106-METJA STANDARD; PRT; 238 AA.  
AC Q57570;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE HYPOTHETICAL PROTEIN MJ0106.  
GN MJ0106.  
OS Methanococcus jannaschii.  
OC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;  
OC Methanococcus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;  
RX MEDLINE; 96337999.  
RA BULT C.J., WHITE O., OLSEN G.J., ZHOU L., FLEISCHMANN R.D.,  
RA SUTTON G.G., BLAKE J.A., FITZGERALD L.M., CLAYTON R.A., GOCAYNE J.D.,  
RA KERLAVAGE A.R., DOUGHERTY B.A., TOMB J.-F., ADAMS M.D., REICH C.I.,  
RA OVERBECK R., KIRKNESS E.F., WEINSTOCK K.G., MERRICK J.M., GLODEK A.,  
RA SCOTT J.L., GEOGHAGEN N.S.M., WEIDMAN J.F., FUHRMANN J.L., NGUYEN D.,  
RA UTTERBACK T.R., KELLEY J.M., PETERSON J.D., SADOW P.W., HANNA M.C.,  
RA COTTON M.D., ROBERTS K.M., HURST M.A., KAINE B.P., BORODOVSKY M.,  
RA KLENK H.-P., FRASER C.M., SMITH H.O., WOESSE C.R., VENTER J.C.;  
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus  
RT jannaschii";  
RL Science 273:1058-1073(1996).  
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).  
CC -!- SIMILARITY: WEAK, TO MJANNASCHII MJ210.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way

Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 829 IPYSVVRG 836  
| | | : | | :  
QY 1 IPYPIVRS 8

RESULT 11  
ID COS\_MOUSE STANDARD; PRT; 1680 AA.  
AC P06684;  
DT 01-JAN-1988 (Rel. 06, Created)  
DT 01-NOV-1990 (Rel. 16, Last sequence update)  
DT 01-FEB-1996 (Rel. 33, Last annotation update)  
DE COMPLEMENT C5 PRECURSOR [CONTAINS: C5A ANAPHYLATOXIN].  
GN C5 OR HC.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
RN [1]  
RN SEQUENCE FROM N.A.  
RX MEDLINE: 90153853.  
RA WETSEL R.A., FLEISCHER D.T., HAVILAND D.L.;  
RT "Deficiency of the murine fifth complement component (C5). A 2-base  
pair gene deletion in a 5'-exon.";  
RL J. Biol. Chem. 265:2435-2440(1990).  
RN [2]  
RN SEQUENCE OF 41-1680 FROM N.A.  
RX MEDLINE: 87185363.  
RA WETSEL R.A., OGATA R.T., TACK B.F.;  
RT "Primary structure of the fifth component of murine complement.";  
RL Biochemistry 26:737-743(1987).  
CC -1- FUNCTION: ACTIVATION OF C5 BY A C5 CONVERTASE INITIATES THE  
SPONTANEOUS ASSEMBLY OF THE LATE COMPLEMENT COMPONENTS, C5-C9,  
INTO THE MEMBRANE ATTACK COMPLEX. C5B HAS A TRANSIENT BINDING SITE  
FOR C6. THE C5B-C6 COMPLEX IS THE FOUNDATION UPON WHICH THE LYtic  
COMPLEX IS ASSEMBLED.  
CC -1- FUNCTION: DERIVED FROM PROTEOLYTIC DEGRADATION OF COMPLEMENT C5,  
C5 ANAPHYLATOXIN IS A MEDIATOR OF LOCAL INFLAMMATORY PROCESS. IT  
INDUCES THE CONTRACTION OF SMOOTH MUSCLE, INCREASES VASCULAR  
PERMEABILITY AND CAUSES HISTAMINE RELEASE FROM MAST CELLS AND  
BASOPHILIC LEUKOCYTES. C5A ALSO STIMULATES THE LOCOMOTION OF  
POLYMPHONUCLEAR LEUKOCYTES (CHEMOKINESIS) AND DIRECT THEIR  
MIGRATION TOWARD SITES OF INFLAMMATION (CHEMOTAXIS).  
CC -1- SUBUNIT: C5 PRECURSOR IS FIRST PROCESSED BY THE REMOVAL OF 4 BASIC  
RESIDUES, FORMING TWO CHAINS, BETA & ALPHA, LINKED BY A DISULFIDE  
BOND. C5 CONVERTASE ACTIVATES C5 BY CLEAVING THE ALPHA CHAIN,  
RELEASING C5A ANAPHYLATOXIN & GENERATING C5B (BETA CHAIN + ALPHA  
CHAIN).  
CC -1- SIMILARITY: TO C3, C4 AND ALPHA-2-MACROGLOBULIN.  
CC -1- SIMILARITY: CONTAINS 1 ANAPHYLATOXIN-LIKE DOMAIN.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
CC EMBL; M35525; AAA37349.1; -;  
CC EMBL; M35526; AAA37348.1; -;  
CC PIR; A27538; A27538.  
CC PIR; A35530; A35530.  
CC HSP; P01031; 1KJS.  
CC MGD; MGI:96031; HC.  
CC PROSITE; PS00477; ALPHA\_2\_MACROGLOBULIN; FALSE\_NEG.  
CC PROSITE; PS01177; ANAPHYLATOXIN\_1; 1.  
CC PROSITE; PS01178; ANAPHYLATOXIN\_2; 1.  
CC PFAM; PF00207; A2M; 1.  
CC Complement pathway; Complement alternate pathway; Glycoprotein;  
KW Plasma; Membrane attack complex; Cytolysis; Inflammatory response;  
KW Signal.  
FT SIGNAL

1 18

Query Match 71.4%; Score 50; DB 1; Length 1680;  
Best Local Similarity 62.5%; Pred. No. 5.12e+00;  
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 833 IPYSVVRG 840  
| | | : | | :  
QY 1 IPYPIVRS 8

RESULT 12  
ID DYHC\_EMBL STANDARD; PRT; 4344 AA.  
AC P45444;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE DYNEIN HEAVY CHAIN, CYTOSOLIC (DYHC).  
GN NUDA.  
OS Emericella nidulans (Aspergillus nidulans).  
OC Eukaryota; Fungi; Ascomycota; Euascomycetes; Plecomycetes;  
OC Eurotiales; Trichocomaceae; Emericella.  
RN [1]  
RN SEQUENCE FROM N.A.  
RX MEDLINE: 94181539.  
RA XIANG X., BECKWITH S.M., MORRIS R.N.;  
RT "Cytoplasmic dynein is involved in nuclear migration in Aspergillus  
nidulans.";  
RL Proc. Natl. Acad. Sci. U.S.A. 91:2100-2104(1994).  
CC -1- FUNCTION: DYNEIN HAS ATPASE ACTIVITY. CYTOPLASMIC DYNEIN ACTS AS A  
MOTOR FOR THE INTRACELLULAR RETROGRADE MOTILITY OF VESICLES AND  
ORGANELLES ALONG MICROTUBULES. REQUIRED TO MAINTAIN UNIFORM  
NUCLEAR DISTRIBUTION IN HYPHAE.  
CC -1- SUBUNIT: CONSISTS OF AT LEAST TWO HEAVY CHAINS AND A NUMBER OF  
INTERMEDIATE AND LIGHT CHAINS.  
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.  
CC -1- SIMILARITY: BELONGS TO THE DYNEIN HEAVY CHAIN FAMILY.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
CC EMBL; U03904; AAA18338.1; -;  
CC PIR; A53489; A53489.  
CC Motor protein; Microtubules; Dynein; ATP-binding;  
KW Heptad repeat pattern.  
FT NP\_BIND 1933 1940  
FT NP\_BIND 2223 2230  
FT NP\_BIND 2592 2599  
FT NP\_BIND 2932 2939  
FT NP\_BIND 4344 4344  
FT SEQUENCE 4344 AA; 492470 MW; 1D75C7EB CRC32;  
Query Match 71.4%; Score 50; DB 1; Length 4344;



AC DT 21-JUL-1986 (Rel. 01, Created)  
DT 01-DEC-1992 (Rel. 24, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE COMPLEMENT C5 PRECURSOR [CONTAINS: C5A ANAPHYLATOXIN].  
GN C5.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eukarya; Primates; Catarrhini; Hominidae; Homo.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 91079575.  
RA HAVILAND D.L., HAVILAND J.C., FLEISCHER D.T., HUNT A., WETSEL R.A.;  
RT "Complete cDNA sequence of human complement pro-C5. Evidence of  
RT truncated transcripts derived from a single copy gene.";  
RL J. Immunol. 146:362-368(1991).  
RN [2]  
RP SEQUENCE OF 412-1676 FROM N.A.  
RX MEDLINE; 88209511.  
RA WETSEL R.A., LEMONS R.S., LEBEAU M.M., BARNUM S.R., NOACK D.,  
RA TACK B.F.;  
RT "Molecular analysis of human complement component C5: localization of  
RT the structural gene to chromosome 9.";  
RL Biochemistry 27:1474-1482(1988).  
RN [3]  
RP SEQUENCE OF 412-902 FROM N.A.  
RX MEDLINE; 85130937.  
RA LUNDWALL A.B., WETSEL R.A., KRISTENSEN T., WHITEHEAD A.S.,  
RA WOODS D.E., OGDEN R.C., COLTEN H.R., TACK B.F.;  
RT "Isolation and sequence analysis of a cDNA clone encoding the fifth  
RT complement component.";  
RL J. Biol. Chem. 260:2108-2112(1985).  
RN [4]  
RP SEQUENCE OF 678-751.  
RX MEDLINE; 79005687.  
RA FERNANDEZ H.N., HUGLI T.E.;  
RT "Primary structural analysis of the polypeptide portion of human C5a  
RT anaphylatoxin. Polypeptide sequence determination and assignment of  
RT the oligosaccharide attachment site in C5a.";  
RL J. Biol. Chem. 253:6955-6964(1978).  
RN [5]  
RP SEQUENCE OF 678-751 FROM N.A.  
RX MEDLINE; 91144547.  
RA BORNISACK J.F., MOLLISON K.W., BUKO A.M., ASHWORTH J.C., HILL H.R.;  
RT "Group B streptococci inactivate complement component C5a by enzymic  
RT cleavage at the C-terminus.";  
RL Biochem. J. 273:635-640(1991).  
RN [6]  
RP STRUCTURE BY NMR OF C5A.  
RX MEDLINE; 88309734.  
RA ZUIDERWEG E.R., MOLLISON K.W., HENKIN J., CARTER G.W.;  
RT "Sequence-specific assignments in the 1H NMR spectrum of the human  
RT inflammatory protein C5a.";  
RL Biochemistry 27:3568-3580(1988).  
RN [7]  
RP STRUCTURE BY NMR OF C5A.  
RX MEDLINE; 89207527.  
RA ZUIDERWEG E.R., NETTESHEIM D.G., MOLLISON K.W., CARTER G.W.;  
RT "Tertiary structure of human complement component C5a in solution  
RT from nuclear magnetic resonance data.";  
RL Biochemistry 28:172-185(1989).  
RN [8]  
RP STRUCTURE BY NMR OF C5A.  
RX MEDLINE; 89274164.  
RA ZUIDERWEG E.R., FESIK S.W.;  
RT "Heteronuclear three-dimensional NMR spectroscopy of the inflammatory  
RT protein C5a.";  
RL Biochemistry 28:2387-2391(1989).  
RN [9]  
RP STRUCTURE BY NMR OF C5A.  
RX MEDLINE; 97160477.  
RA ZHANG X., BOJAR W., GALAKATOS N., GONNELLA N.C.;  
RT "Solution structure of a unique C5a semi-synthetic antagonistist:  
RT Implications in receptor binding.";  
Protein Sci. 6:65-72(1997).  
RN [10]  
RP STRUCTURE BY NMR OF C5A.  
RX MEDLINE; 97332508.  
RA ZHANG X., BOJAR W., TOTH M.J., WENNOGLE L., GONNELLA N.C.;  
RT "Structural definition of the C5a C terminus by two-dimensional  
RT nuclear magnetic resonance spectroscopy.";  
RL Proteins 28:261-267(1997).  
RN [11]  
RP FUNCTION: ACTIVATION OF C5 BY A C5 CONVERTASE INITIATES THE  
CC SPONTANEOUS ASSEMBLY OF THE LATE COMPLEMENT COMPONENTS, C5-C9,  
CC INTO THE MEMBRANE ATTACK COMPLEX. C5B HAS A TRANSIENT BINDING SITE  
CC FOR C6. THE C5B-C6 COMPLEX IS THE FOUNDATION UPON WHICH THE LYtic  
CC COMPLEX IS ASSEMBLED.  
CC -1- FUNCTION: DERIVED FROM PROTEOLYTIC DEGRADATION OF COMPLEMENT C5,  
CC C5 ANAPHYLATOXIN IS A MEDIATOR OF LOCAL INFLAMMATORY PROCESS. IT  
CC INDUCES THE CONTRACTION OF SMOOTH MUSCLE, INCREASES VASCULAR  
CC PERMEABILITY AND CAUSES HISTAMINE RELEASE FROM MAST CELLS AND  
CC BASOPHILIC LEUKOCYTES. C5A ALSO STIMULATES THE LOCOMOTION OF  
CC POLYMORPHONUCLEAR LEUKOCYTES (CHEMOKINESIS) AND DIRECT THEIR  
CC MIGRATION TOWARD SITES OF INFLAMMATION (CHEMOTAXIS).  
CC -1- SUBUNIT: C5 PRECURSOR IS FIRST PROCESSED BY THE REMOVAL OF 4 BASIC  
CC RESIDUES, FORMING TWO CHAINS, BETA & ALPHA, LINKED BY A DISULFIDE  
CC BOND. C5 CONVERTASE ACTIVATES C5 BY CLEAVING THE ALPHA CHAIN,  
CC RELEASING C5A ANAPHYLATOXIN & GENERATING C5B (BETA CHAIN + ALPHA  
CC CHAIN).  
CC -1- SIMILARITY: TO C3, C4 AND ALPHA-2-MACROGLOBULIN.  
CC -1- SIMILARITY: CONTAINS 1 ANAPHYLATOXIN-LIKE DOMAIN.  
CC -1- CAUTION: REF.3 SEQUENCE DIFFERS FROM THAT SHOWN FROM POSITION 855  
CC ONWARD DUE TO THE PRESENCE OF AN ALU REPEAT.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [licensed@isb-sib.ch](mailto:licensed@isb-sib.ch)).  
CC -----  
DR EMBL; M57729; AAA51925.1; -  
DR EMBL; M65134; AAA51856.1; -  
DR PIR; A40075; C5HU.  
DR PIR; S15121; S15121.  
DR PDB; 1KJS; 15-MAY-97.  
DR PDB; 1CFA; 17-SEP-97.  
DR MIM; 120900; -  
DR PROSITE; PS00477; ALPHA-2-MACROGLOBULIN; FALSE\_NEG.  
DR PROSITE; PS01177; ANAPHYLATOXIN\_1; 1.  
DR PROSITE; PS01178; ANAPHYLATOXIN\_2; 1.  
DR PFAM; PF00207; A2M; 1.  
DR KW Complement pathway; Complement alternate pathway; Glycoprotein;  
KW Plasma; Membrane attack complex; Cytolysis; Inflammatory response;  
KW Signal; Polymorphism; 3D-structure.  
FT SIGNAL 1 18  
FT CHAIN 19 673 COMPLEMENT C5 BETA CHAIN.  
FT PROPEP 674 677  
FT CHAIN 678 1676 COMPLEMENT C5 ALPHA CHAIN.  
FT PEPTIDE 678 751 C5A ANAPHYLATOXIN.  
FT CHAIN 752 1676 C5B (ALPHA').  
FT DOMAIN 698 732 ANAPHYLATOXIN-LIKE.  
FT DISULFID 698 724  
FT DISULFID 699 731  
FT DISULFID 711 732  
FT CARBOHYD 741 741  
FT CARBOHYD 911 911 POTENTIAL.  
FT CARBOHYD 1115 1115 POTENTIAL.  
FT CARBOHYD 1630 1630 POTENTIAL.  
FT VARIANT 518 518 F -> S.  
SQ SEQUENCE 1676 AA; 188331 MW; 9D5C6E59 CRC32;  
/FTID=VAR\_001996.  
Query Match 71.48; Score 50; DB 1; Length 1676;  
Best Local Similarity 62.5%; Pred. No. 5.12e+00;

```
CC EMBL; Z50019; CAA90322.1; -
DR PROSITE; PS00677; DAO; 1.
DR PFAM; PF01266; DAO; 1.
KW Oxidoreductase; Flavoprotein; FAD.
FT NP_BIND 4 18 FAD (ADP PART) (POTENTIAL).
FT ACT_SITE 243 243 BY SIMILARITY.
FT ACT_SITE 324 324 BY SIMILARITY.
SQ SEQUENCE 356 AA; 39301 MW; BA069642 CRC32;

Query Match 71.4%; Score 50; DB 1; Length 356;
Best Local Similarity 55.6%; Pred. No. 5.12e+00;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 66 VSPILREL 74
QY 1 IPIPIVRS 9

RESULT 9
ID RNE_ECOLI STANDARD; PRT; 1061 AA.
AC P21513; P77591;
DT 01-MAY-1991 (Rel. 18, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE RIBONUCLEASE E (EC 3.1.4.-) (RNASE E).
GN RNE OR AMS OR HMP1.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE; 97426617.
RA BLATNER F.R., PLUNKETT G. III, BLOCH C.A., PERNA N.T., BURLAND V.,
RA RILEY M., COLLADO-VIDES J., GLASNER F.D., RODE C.K., MAYHEW G.F.,
RA GREGOR J., DAVIS N.W., KIRKPATRICK H.A., GOEDEN M.A., ROSE D.J.,
RA MAU B., SHAO Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RX MEDLINE; 97061202.
RA OSHIMA T., AIBA H., BABA T., FUJITA K., HAYASHI K., HONJO A.,
RA IKEMOTO K., INADA T., ITOH T., KAJIHARA M., KANAI K., KASHIMOTO K.,
RA KIMURA S., KITAGAWA M., MAKINO K., MASUDA S., MIKI T., MIZOBUCHI K.,
RA MORI H., MOTOMURA K., NAKAMURA Y., NASHIMOTO H., NISHIO Y., SAITO N.,
RA SAMPEI G., SERI Y., TAGAMI H., TAKEMOTO K., WADA C., YAMAMOTO Y.,
RA YANO M., HORIUCHI T.;
RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 12.7-28.0 min region on the linkage map.";
RL DNA Res. 3:137-155(1996).
RN [3]
RP SEQUENCE OF 1-1025 FROM N.A.
RC STRAIN-K12;
RX MEDLINE; 93078265.
RA CASAREGOLA S., JACQ A., LAOUDJ D., MCGURK G., MARGARSON S.,
RA TEMPETE M., NORRIS V., HOLLAND I.B.;
RT "Cloning and analysis of the entire Escherichia coli ams gene. ams is
RT identical to hmp1 and encodes a 114 kDa protein that migrates as a
RT 180 kDa protein.";
RL J. Mol. Biol. 228:30-40(1992).
RN [4]
RP SEQUENCE OF 1-844 FROM N.A.
RC STRAIN-K12;
RX MEDLINE; 91131576.
RA CLAYERIE-MARTIN F., DIAZ-TORRES M., YANCEY S.D., KUSHNER S.R.;
RT "Analysis of the altered mRNA stability (ams) gene from Escherichia
RT coli. Nucleotide sequence, transcriptional analysis, and homology of
RT its product to MRP3, a mitochondrial ribosomal protein from
RT Neurospora crassa.";
RL J. Biol. Chem. 266:2843-2851(1991).
```

```

RN [5]
RP PARTIAL SEQUENCE FROM N.A., AND SEQUENCE OF 1-27.
RC STRAIN-K12;
RX MEDLINE; 91187608.
RA CHAUHAN A.K., MICZAK A., TARASEVICIENE L., APIRION D.;
RT "Sequencing and expression of the rne gene of Escherichia coli.";
RL Nucleic Acids Res. 19:125-129(1991).
RN [6]
RP SEQUENCE OF 844-1061 FROM N.A., AND CHARACTERIZATION.
RC STRAIN-K12;
RX MEDLINE; 94022304.
RA CORMACK R.S., GENEVAUX J.L., MACKIE G.A.;
RT "RNase E activity is conferred by a single polypeptide:
RT overexpression, purification, and properties of the ams/rne/hmp1 gene
RT product.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:9006-9010(1993).
CC -!- FUNCTION: THIS PROTEIN MATURES 5S RNA FROM ITS PRECURSORS FROM
CC ALL THE RNA GENES. IT ALSO CLEAVES RNA I, A MOLECULE THAT
CC CONTROLS THE REPLICATION OF COLE1 PLASMID DNA. IT IS THE MAJOR
CC ENDORIBONUCLEASE PARTICIPATING IN MRNA TURNOVER IN E.COLI.
CC -!- SUBUNIT: ORGANIZED INTO A STRUCTURE (PROCESSOME OR RNA
CC DEGRADOSOME) CONTAINING A NUMBER OF RNA-PROCESSING ENZYMES.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -!- SIMILARITY: BELONGS TO THE RNE FAMILY.
CC -!- CAUTION: REF.3 SEQUENCE DIFFERS FROM THAT SHOWN FROM POSITION 1003
CC ONWARD AND IS SHORTER (1025 AA) DUE TO A FRAMESHIFT.
CC -!- CAUTION: REF.4 SEQUENCE DIFFERS FROM THAT SHOWN IN THE C-TERMINUS
CC AND IS SHORTER (815 AA) DUE TO A FRAMESHIFT.
CC -!- CAUTION: REF.5 SEQUENCE WAS ALSO INCORRECT IN MANY POSITIONS DUE
CC TO FRAMESHIFTS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE000209; AAC74168.1; -
CC EMBL; D90744; BAA35893.1; -
CC EMBL; X67470; CAA47818.1; ALT_FRAME.
CC EMBL; M62747; AAA23443.1; ALT_FRAME.
CC EMBL; X54309; CAA38206.1; ALT_FRAME.
CC EMBL; L23942; AAA03347.1; -
CC PIR; JG0009; JG0009.
CC PIR; A23747; A23747.
CC PIR; S25116; S25116.
CC PIR; S27311; S27311.
CC HSSP; P05055; 1SRO.
CC ECGENE; EG10859; RNE.
CC PFAM; PF00575; S1; 1.
KW Hydrolase; Nuclease; Endonuclease; RNA-binding.
FT CONFLICT 390 390 Q -> H (IN REF. 4).
FT CONFLICT 487 487 V -> L (IN REF. 1 AND 2).
FT CONFLICT 564 564 A -> R (IN REF. 3).
FT CONFLICT 784 784 N -> K (IN REF. 3).
FT CONFLICT 838 838 A -> R (IN REF. 4).
FT CONFLICT 905 905 P -> R (IN REF. 3).
FT CONFLICT 1048 1048 H -> R (IN REF. 6).
SQ SEQUENCE 1061 AA; 118182 MW; 2CE7D241 CRC32;

Query Match 71.4%; Score 50; DB 1; Length 1061;
Best Local Similarity 75.0%; Pred. No. 5.12e+00;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 844 IRYPIVRP 851
QY 1 IPIPIVRS 8

RESULT 10
ID COS_HUMAN STANDARD; PRT; 1676 AA.
```

```

RESULT 6
ID   ARG1ARATH          STANDARD;          PRT;      342 AA.
AC   P46637;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   ARGINASE (EC 3.5.3.1).
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC   euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC   core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
OC   Arabidopsids.
[1]
RN   SEQUENCE FROM N.A.
RC   STRAIN-CV, LANDSBERG ERECTA;
RX   MEDLINE; 95288383.
RA   KUMPELMAN P.M., FREYERMUTH S.K., CANNON J.F., FINK G.R.,
RA   POLACCO J.C.;
RT   "Nucleotide sequence of Arabidopsis thaliana arginase expressed in
RL   yeast.";
RL   Plant Physiol. 107:1479-1480(1995).
CC   -1- CATALYTIC ACTIVITY: L-ARGININE + H(2)O -> L-ORNITHINE + UREA.
CC   -1- COFACTOR: MANGANESE (BY SIMILARITY).
CC   -1- PATHWAY: FIRST STEP IN ARGinine DEGRADATION.
CC   -1- SIMILARITY: BELONGS TO THE ARGINASE FAMILY.
-----
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
-----
DR   EMBL; U15019; AA85816.1; -
DR   PROSITE; PS00147; ARGINASE_1; 1.
DR   PROSITE; PS00148; ARGINASE_2; 1.
DR   PROSITE; PS01053; ARGINASE_3; 1.
DR   PFAM; PF00491; arginase; 1
KW   Hydrolase; Arginine metabolism; Manganese.
FT   METAL 161
FT   METAL 161 MANGANESE 1 (BY SIMILARITY).
FT   METAL 185
FT   METAL 187 MANGANESE 1 AND 2 (BY SIMILARITY).
FT   METAL 187 MANGANESE 2 (BY SIMILARITY).
FT   METAL 189
FT   METAL 189 MANGANESE 1 (BY SIMILARITY).
FT   METAL 270
FT   METAL 270 MANGANESE 1 AND 2 (BY SIMILARITY).
FT   METAL 272
FT   METAL 272 MANGANESE 2 (BY SIMILARITY).
SQ   SEQUENCE 342 AA; 37344 MW; 9640021A CRC32;
Query Match 74.3%; Score 52; DB 1; Length 342;
Best Local Similarity 55.6%; Pred. No. 1.87e+00;
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 163 ISPPVVRV 171
|:|:|:|:
QY 1 IPYPIVRS 9

RESULT 7
ID   THII1METJA          STANDARD;          PRT;      381 AA.
AC   Q58341;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   PROBABLE THIAMINE BIOSYNTHESIS PROTEIN THII.
GN   THII OR MJ0931.
OS   Methanococcus jannaschii.
OC   Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
OC   Methanococcus.
[1]
RN   SEQUENCE FROM N.A.
RC   STRAIN-JAL-1 / DSM 2661 / ATCC 43067;

RX   MEDLINE; 96337799.
RA   BULT C.J., WHITE O., OLSEN G.J., ZHOU L., FLEISCHMANN R.D.,
RA   SUTTON G.G., BLAKE J.A., FITZGERALD L.M., CLAYTON R.A., GOCAYNE J.D.,
RA   KERLAVAGE A.R., DOUGHERTY B.A., TOMB J.-F., ADAMS M.D., REICH C.I.,
RA   OVERBEK R., KIRKNESS E.F., WEINSTOCK K.G., MERRICK J.M., GLODEK A.,
RA   SCOTT J.L., GEORHAGEN N.S.M., WEIDMAN J.F., FUHRMANN J.L., NGUYEN D.,
RA   UTTERBACK T.R., KELLEY J.M., PETERSON J.D., SADOW P.W., HANNA M.C.,
RA   COTTON M.D., ROBERTS K.M., HURST M.A., KAINE B.P., BORODOVSKY M.,
RA   KLENK H.-P., FRASER C.M., SMITH H.O., WOESE C.R., VENTER J.C.;
RT   "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT   jannaschii.";
RL   Science 273:1058-1073(1996).
CC   -1- FUNCTION: REQUIRED FOR THE SYNTHESIS OF THE THIAZOLE MOIETY (BY
CC   SIMILARITY).
CC   -1- PATHWAY: THIAMINE BIOSYNTHESIS.
CC   -1- SUBCELLULAR LOCATION: CYTOPLASMIC (POTENTIAL).
CC   -1- SIMILARITY: BELONGS TO THE THII FAMILY.
-----
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
-----
DR   EMBL; U67536; AAB98933.1; -
DR   TIGR; MJ0931; -
KW   Thiamine biosynthesis.
SQ   SEQUENCE 381 AA; 43436 MW; 853CF1A9 CRC32;
Query Match 72.9%; Score 51; DB 1; Length 381;
Best Local Similarity 66.7%; Pred. No. 3.11e+00;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 319 INYPILRPL 327
|:|:|:|:
QY 1 IPYPIVRS 9

RESULT 8
ID   OXDA1TRIVR          STANDARD;          PRT;      356 AA.
AC   Q99042;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   D-AMINO ACID OXIDASE (EC 1.4.3.3) (DAMO) (DAO).
GN   DAO1.
OS   Trigonopsis variabilis.
OC   Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
OC   Candidaceae; Trigonopsis.
[1]
RN   SEQUENCE FROM N.A.
RC   STRAIN-CBS 4095;
RX   MEDLINE; 98095789.
RA   GONZALEZ F.J., MONTES J., MARTIN F., LOPEZ M.C., FERMINAN E.,
RA   CATALAN J., GALAN M.A., DOMINGUEZ A.;
RT   "Molecular cloning of TvDAO1, a gene encoding a D-amino acid oxidase
RT   from Trigonopsis variabilis and its expression in Saccharomyces
RT   cerevisiae and Kluyveromyces fragilis.";
RL   Yeast 13:1399-1408(1997).
CC   -1- CATALYTIC ACTIVITY: A D-AMINO ACID + H(2)O + O(2) -> A 2-OXO-ACID +
CC   NH(3) + H(2)O(2).
CC   -1- COFACTOR: FAD.
CC   -1- SIMILARITY: BELONGS TO THE DAMOX/DASOX FAMILY.
-----
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
```



```
QY 1 IPYPIVRS 9
RESULT 2
ID RAS_GEOCY STANDARD; PRT; 209 AA.
AC P24498;
DT 01-MAR-1992 (Rel. 21, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE RAS-LIKE PROTEIN.
OS Geodia cydonium (Sponge).
OC Eukaryota; Metazoa; Porifera; Demospongiae; Tetractinomorpha;
OC Astrophorida; Geodidae; Geodia.
RN [1]
SEQUENCE FROM N.A.
RP MEDLINE; 91006138.
RA ROBITZKI A., SCHROEDER H.C., UGARKOVIC D., KUCHINO Y., KURELEC B.,
RA GAMULIN V., MUELLER W.E.G.; Phosphorylation of the 23-26-kDa ras
RT protein in the sponge Geodia cydonium."
RL Eur. J. Biochem. 192:499-506(1990).
CC -1- FUNCTION: THIS PROTEIN IS ACTIVATED BY THE INSULIN/INSULIN
CC (INSULIN-LIKE)-RECEPTOR SYSTEM. THIS TRANSITION ENABLES THE RAS
CC PROTEIN TO INTERACT WITH THE LECTIN-RECEPTOR/LECTIN COMPLEX, A
CC PROCESS WHICH ULTIMATELY LEAD TO AN INITIATION OF AN INTRA-
CC CELLULAR SIGNAL-TRANSDUCTION CHAIN.
CC -1- ENZYME REGULATION: ALTERNATE BETWEEN AN INACTIVE FORM BOUND TO GDP
CC AND AN ACTIVE FORM BOUND TO GTP. ACTIVATED BY A GUANINE
CC NUCLEOTIDE-EXCHANGE FACTOR (GEF) AND INACTIVATED BY A GTPASE-
CC ACTIVATING PROTEIN (GAP).
CC -1- PTM: PHOSPHORYLATED IN THE PRESENCE OF INSULIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: M30929; NOT_ANNOTATED_CDS.
CC FIR; S13179; S13179.
CC DR HSSP; P01112; IPLJ.
CC DR PFAM; PF00071; ras; 1.
CC KW GTP-binding; Prenylation; Lipoprotein; Phosphorylation.
CC FT NP_BIND 10 17 GTP (BY SIMILARITY).
CC FT NP_BIND 79 83 GTP (BY SIMILARITY).
CC FT NP_BIND 140 143 GTP (BY SIMILARITY).
CC FT DOMAIN 55 63 EFFECTOR REGION (BY SIMILARITY).
CC FT MOD_RES 58 58 PHOSPHORYLATION (POTENTIAL).
CC FT LIPID 206 206 GERANYL-GERANYL (BY SIMILARITY).
CC SQ SEQUENCE 209 AA; 23854 MW; E07739EF CRC32;
Query Match 75.7%; Score 53; DB 1; Length 209;
Best Local Similarity 66.7%; Pred. No. 1.12e+00;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Db 178 IPYSLVREL 186
QY 1 IPYPIVRS 9
RESULT 3
ID ARG1_SOYBN STANDARD; PRT; 350 AA.
AC Q49046;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE ARGINASE (EC 3.5.3.1).
OS Agl.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae;
Glycine.
RN [1]
SEQUENCE FROM N.A.
RP STRAIN-CV. WILLIAMS 82;
RA GOLDRAIJ A., COELLO P., POLACCO J.C.;
RT "Nucleotide sequence of a cDNA encoding a soybean seedling axes
RT arginase."
RL (In) Plant Gene Register PGR98-016.
CC -1- CATALYTIC ACTIVITY: L-ARGININE + H(2)O = L-ORNITHINE + UREA.
CC -1- COFACTOR: MANGANESE (BY SIMILARITY).
CC -1- PATHWAY: FIRST STEP IN ARGININE DEGRADATION.
CC -1- SIMILARITY: BELONGS TO THE ARGINASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AF035671; AAC04613.1;
CC DR PROSITE; PS00147; ARGINASE_1; FALSE_NEG.
CC DR PROSITE; PS00148; ARGINASE_2; 1.
CC DR PROSITE; PS01053; ARGINASE_3; FALSE_NEG.
CC DR PFAM; PF00491; arginase; 1.
CC KW Hydrolase; Arginine metabolism; Manganese.
CC FT METAL 193 193 MANGANESE 1 AND 2 (BY SIMILARITY).
CC FT METAL 195 195 MANGANESE 2 (BY SIMILARITY).
CC FT METAL 197 197 MANGANESE 1 (BY SIMILARITY).
CC FT METAL 278 278 MANGANESE 1 AND 2 (BY SIMILARITY).
CC FT METAL 280 280 MANGANESE 2 (BY SIMILARITY).
CC SQ SEQUENCE 350 AA; 38610 MW; 8D66800E CRC32;
Query Match 75.7%; Score 53; DB 1; Length 350;
Best Local Similarity 55.6%; Pred. No. 1.12e+00;
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
Db 171 ISVPVRAI 179
QY 1 IPYPIVRS 9
RESULT 4
ID TTDI_ECOLI STANDARD; PRT; 487 AA.
AC P39414; Q46870;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE PUTATIVE TARTRATE CARRIER (TARTRATE TRANSPORTER) (TARTRATE/SUCCINATE
DE ANTIporter).
GN YGJE.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
RN [1]
SEQUENCE FROM N.A.
RP STRAIN-K12 / MG1655;
RX MEDLINE; 97426617.
RA BLATTNER F.R., PLUNKETT G. III, BLOCH C.A., PERNA N.T., BURLAND V.,
RA RILEY M., COLLADO-VIDES J., GLASNER F.D., RODE C.K., MAYHEW G.F.,
RA GREGOR J., DAVIS N.W., KIRKPATRICK H.A., GOEDEN M.A., ROSE D.J.,
RA MAU B., SHAO Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
RN [2]
SEQUENCE FROM N.A.
RP MEDLINE; 87248073.
RA NESIN M., LUPSKI J.R., SVEC P., GODSON G.N.;
RT "Possible new genes as revealed by molecular analysis of a 5-kb
RT Escherichia coli chromosomal region 5' to the rpsU-dnaG-rpoD
RT macromolecular-synthesis operon."
RN [2]
```

\*\*\*\*\*  
W P S R E H  
\*\*\*\*\*  
(TM)  
\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

Mpsrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Apr 14 23:49:52 2000: Maspar time 6.06 Seconds  
Tabular output not generated. 44.371 Million cell updates/sec

Title: >US-08-452-843-9  
Description: (1-9) from US08452843.pep  
Perfect Score: 70  
Sequence: 1 IYPPIVRSLS 9  
Scoring table: PAM 150  
Gap 15  
Searched: 82229 seqs, 29864866 residues  
Post-processing: Minimum Match 0%  
Listing first 45 summaries  
Database: swiss-prot38  
1:swissprot  
Statistics: Mean 24.809; Variance 27.222; scale 0.911

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	56	80.0	340	1	RPC2_SCHPO PROBABLE ACTIVATOR 1 4	2.30e+01
2	53	75.7	209	1	RAS_GOCY RAS-LIKE PROTEIN	1.12e+00
3	53	75.7	350	1	ARGI_SOYBN ARGINASE (EC 3.5.3.1).	1.12e+00
4	53	75.7	487	1	TIDT_ECOLI PUTATIVE TARTRATE CARR	1.12e+00
5	52	74.3	133	1	SY21_MOUSE SMALL INDUCIBLE CYTOKI	1.87e+00
6	52	74.3	342	1	ARGI_ARATH ARGINASE (EC 3.5.3.1).	1.87e+00
7	51	72.9	381	1	THIL_METUA PROBABLE THIAMINE BIOS	3.11e+00
8	50	71.4	356	1	OXDA_TRIVR D-AMINO ACID OXIDASE (	5.12e+00
9	50	71.4	1061	1	RNE_ECOLI RIBONUCLEASE E (EC 3.1	5.12e+00
10	50	71.4	1676	1	COS_HUMAN COMPLEMENT C5 PRECURSO	5.12e+00
11	50	71.4	1680	1	COS_MOUSE COMPLEMENT C5 PRECURSO	5.12e+00
12	50	71.4	4344	1	DYHC_EMENT DYNEIN HEAVY CHAIN, CY	5.12e+00
13	50	71.4	4349	1	DYHC_FUSSO DYNEIN HEAVY CHAIN, CY	5.12e+00
14	50	71.4	4367	1	DYHC_NEUCR DYNEIN HEAVY CHAIN, CY	5.12e+00
15	49	70.0	238	1	Y106_METUA HYPOTHETICAL PROTEIN M	8.36e+00
16	49	70.0	298	1	LACF_AGRRD LACTOSE TRANSPORT SYST	8.36e+00
17	48	68.6	78	1	Y080_METUA HYPOTHETICAL PROTEIN M	1.35e+01
18	48	68.6	305	1	LIGD_PSEPA C ALPHA-DEHYDROGENASE	1.35e+01
19	48	68.6	356	1	YDGC_SCHPO HYPOTHETICAL 41.3 KD P	1.35e+01
20	48	68.6	440	1	NAM1_YEAST NAM1 PROTEIN PRECURSOR	1.35e+01
21	48	68.6	531	1	TRPC_PHYPR TRYPTOPHAN BIOSYNTHESI	1.35e+01
22	48	68.6	880	1	RPAL_SULAC DNA-DIRECTED RNA POLYM	1.35e+01
23	47	67.1	230	1	DAG_ANTMA DAG PROTEIN, CHLOROPLA	2.18e+01

24	47	67.1	268	1	CPCE_SYN2	PHYCOCYANOBILIN LYASE	2.18e+01
25	47	67.1	272	1	CYNT_SYN7	CARBONIC ANHYDRASE (EC	2.18e+01
26	47	67.1	311	1	HTRB_HABIN	LIPID A BIOSYNTHESIS L	2.18e+01
27	47	67.1	356	1	VP39_NPVLD	MAJOR CAPSID PROTEIN.	2.18e+01
28	47	67.1	434	1	YUGS_BACSU	HYPOTHETICAL 49.5 KD P	2.18e+01
29	47	67.1	634	1	HS71_LEIMA	HEAT SHOCK 70-RELATED	2.18e+01
30	47	67.1	675	1	HS7M_PEA	HEAT SHOCK 70 KD PROTE	2.18e+01
31	47	67.1	718	1	PLSB_CABEL	PROBABLE GLYCEROL-3-PH	2.18e+01
32	47	67.1	1056	1	YNN2_YEAST	HYPOTHETICAL 119.3 KD	2.18e+01
33	47	67.1	1082	1	RBL2_HUMAN	RETINOBLASTOMA-LIKE PR	2.18e+01
34	47	67.1	1230	1	UGS4_SOLTU	SOLUBLE GLYCOCEN (STAR	2.18e+01
35	47	67.1	1679	1	Y109_YEAST	HYPOTHETICAL 195.1 KD	2.18e+01
36	46	65.7	98	1	RL2_METJA	50S RIBOSOMAL PROTEIN	3.47e+01
37	46	65.7	284	1	Y309_METJA	HYPOTHETICAL PROTEIN M	3.47e+01
38	46	65.7	291	1	BACH_NATPH	HALORHODOPSIN (HR).	3.47e+01
39	46	65.7	369	1	RF2_RICPR	PEPTIDE CHAIN RELEASE	3.47e+01
40	46	65.7	607	1	G6PI_TRYBB	GLUCOSE-6-PHOSPHATE IS	3.47e+01
41	46	65.7	682	1	HS7M_SOLTU	HEAT SHOCK 70 KD PROTE	3.47e+01
42	46	65.7	1330	1	VCAP_PPRVS	MAJOR CAPSID PROTEIN (	3.47e+01
43	46	65.7	1376	1	VCAP_HSVEB	MAJOR CAPSID PROTEIN (	3.47e+01
44	46	65.7	1441	1	VGLM_BUNL7	M POLYPROTEIN PRECURSO	3.47e+01
45	46	65.7	1453	1	VP15_YEAST	PROTEIN KINASE VPS15 (	3.47e+01

ALIGNMENTS

RESULT 1  
ID RPC2\_SCHPO STANDARD; PRT; 340 AA.  
AC Q09843;  
DT 01-FEB-1996 (Rel. 33, Created)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)  
DT 01-FEB-1996 (Rel. 33, Last annotation update)  
DE PROBABLE ACTIVATOR 1 41 KD SUBUNIT (REPLICATION FACTOR C 41 KD  
DE SUBUNIT).  
GN SPAC23D3.02  
OS Schizosaccharomyces pombe (fission yeast).  
OC Eukaryota; Fungi; Ascomycota; Archiascomycetes;  
OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
OC Schizosaccharomycetes.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-972;  
RA NBLETT D., HARRIS D., BARRELL B.G., RAJANDREAM M.A., WALSH S.V.;  
RL Submitted (OCT-1995) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: THE ELONGATION OF PRIME DNA TEMPLATES BY DNA POLYMERASE  
DELTA AND EPSILON REQUIRES THE ACTION OF THE ACCESSORY PROTEINS  
CC PROLIFERATING CELL NUCLEAR ANTIGEN (PCNA) AND ACTIVATOR 1. THE  
CC 41 KD SUBUNIT BINDS ATP AND TO SINGLE-STRANDED DNA  
CC (BY SIMILARITY).  
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).  
CC -1- SIMILARITY: BELONGS TO THE ACTIVATOR 1 36 TO 40 KD SUBUNITS  
CC FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; 264354; CAA91237.1;  
DR Hypothetical protein; DNA replication; ATP-binding; Nuclear protein;  
KW DNA-binding.  
FT NP\_BIND 59 66 ATP (POTENTIAL).  
SQ SEQUENCE 340 AA; 37876 MW; FB518443 CRC32;  
Query Match 80.0%; Score 56; DB 1; Length 340;  
Best Local Similarity 66.7%; Pred. No. 2.30e+01;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
Db 249 VPIINIRSL 257  
::: |||

**THIS PAGE BLANK (USPTO)**

```
RESULT 14
ENTRY
TITLE      A53489      #type complete
ORGANISM    dynein heavy chain, cytosolic - Emericella nidulans
DATE        #formal_name Emericella nidulans, Aspergillus nidulans
            02-Jun-1994 #sequence_revision 02-Jun-1994 #text_change
            02-Jul-1998
ACCESSIONS  A53489
REFERENCE    Xiang, X.; Beckwith, S.M.; Morris, N.R.
#authors    Proc. Natl. Acad. Sci. U.S.A. (1994) 91:2100-2104
#journal     Cytoplasmic dynein is involved in nuclear migration in
#title       Aspergillus nidulans.
#cross-references MUD:94181539
#accession    A53489
##status      preliminary
##molecule_type DNA
##residues    1-4344 #label XIA
##cross-references GB:U03904; NID:9451538; PID:9451539
SUMMARY     #length 4344 #molecular-weight 492476 #checksum 8396

Query Match      71.4%; Score 50; DB 2; Length 4344;
Best Local Similarity 75.0%; Pred. No. 1.43e+01;
Matches          6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 394 PYPKRAL 401
    |||||
Qy 2 PYPVRSLS 9

RESULT 15
ENTRY
TITLE      B54802      #type complete
ORGANISM    dynein heavy chain, cytosolic - Neurospora crassa
DATE        #formal_name Neurospora crassa
            23-Mar-1995 #sequence_revision 05-Apr-1995 #text_change
            02-Jul-1998
ACCESSIONS  B54802
REFERENCE    Planmann, M.; Minke, P.F.; Tinsley, J.H.; Bruno, K.S.
#authors    J. Cell Biol. (1994) 127:139-149
#journal     Cytoplasmic dynein and actin-related protein Arp1 are
#title       required for normal nuclear distribution in filamentous
            fungi.
#accession    B54802
##status      preliminary
##molecule_type DNA
##residues    1-4367 #label PLA
##cross-references GB:L31504; NID:9473489; PID:9473490
GENETICS
#introns     104/1; 4205/3
SUMMARY     #length 4367 #molecular-weight 495574 #checksum 8268

Query Match      71.4%; Score 50; DB 2; Length 4367;
Best Local Similarity 75.0%; Pred. No. 1.43e+01;
Matches          6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 404 PYPKRAL 411
    |||||
Qy 2 PYPVRSLS 9

Search completed: Fri Apr 14 23:49:34 2000
Job time : 11 secs.
```



anaphylatoxin from the amino end of the alpha chain, generating C5b (beta and alpha' chains).

COMMENT Activation of C5 initiates the spontaneous assembly of the late complement components, C5-C9, into the membrane attack complex. C5b has a transient binding site for C6. The C5b-C6 complex is the foundation upon which the membrane attack complex is assembled.

COMMENT C5a has potent spasmogenic and chemotactic activity.

GENETICS

#gene GDB:C5

#cross-references GDB:119734; OMIM:120900

#map\_position 9q34.1-9q34.1

CLASSIFICATION #superfamily alpha-2-macroglobulin

KEYWORDS complement alternate pathway; complement pathway; cytolysis; glycoprotein; inflammation; membrane attack complex; plasma

FEATURE

1-18 #domain signal sequence #status predicted #label SIG\

19-673,678-1676 #product complement C5 #status predicted #label MAT\

19-673,752-1676 #product C5b #status predicted #label C5B\

19-673 #product complement C5 and C5b beta chain #status predicted #label C5BB\

678-1676 #product complement C5 alpha chain #status predicted #label C5A\

678-751 #product C5a anaphylatoxin #status experimental #label C5T\

752-1676 #product C5b alpha' chain #status predicted #label C5BA\

567-810,634-669,698-724,695-731,711-732,866-1527,1101-1159,1375-1505,1405-1474,1520-1525,1532-1606,1553-1676,1654-1657

751-752 #disulfide\_bonds #status predicted\

#binding\_site carbohydrate (Asn) (covalent) #status experimental\

#cleavage\_site Arg-Leu (C5 convertase) #status experimental\

#binding\_site carbohydrate (Asn) (covalent) #status predicted

911.1115,1630 #length 1676 #molecular-weight 188330 #checksum 3858

SUMMARY

Query Match 71.4%; Score 50; DB 1; Length 1676;

Best Local Similarity 62.5%; Pred. No. 1.43e+01;

Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 829 IPYSVVRG 836

QY 1 IPYPVIRS 8

RESULT 13

ENTRY C5MS #type complete

TITLE Complement C5 precursor - mouse

CONTAINS C5a anaphylatoxin; C5b

ORGANISM #formal\_name Mus musculus #common\_name house mouse

DATE 19-Nov-1988 #sequence\_revision 15-Oct-1994 #text\_change 24-Oct-1997

ACCESSIONS A35530; A27538; A40429

REFERENCE A35530

#authors Wetsel, R.A.; Fleischer, D.T.; Haviland, D.L.

#journal J. Biol. Chem. (1990) 265:2435-2440

#title Deficiency of the murine fifth complement component (C5). A 2-base pair gene deletion in a 5'-exon.

#cross-references MUID:90153853

#accession A35530

#molecule\_type mRNA

#residues 1-215, 'L' #label WET

#cross-references GDB:M35526; GB:J05234; NID:g192302; PID:g309123

REFERENCE A27538

#authors Wetsel, R.A.; Ogata, R.T.; Tack, B.F.

Biochemistry (1987) 26:737-743

#journal Primary structure of the fifth component of murine complement.

#cross-references MUID:87185363

#accession A27538

#molecule\_type mRNA

#residues 'PGL', 44-1680 #label WET2

REFERENCE A40429

#authors Haviland, D.L.; Haviland, J.C.; Fleischer, D.T.; Wetsel, R.A.

#journal J. Biol. Chem. (1991) 266:11818-11825

#title Structure of the murine fifth complement component (C5) gene. A large, highly interrupted gene with a variant donor splice site and organizational homology with the third and fourth complement component genes.

#cross-references MUID:91268053

#accession A40429

#molecule\_type DNA

#residues 1-15 #label HAV

#cross-references GB:M64852

COMMENT Complement C5 contains two disulfide-linked chains, formed by removal of four basic residues. C5 convertase releases C5a anaphylatoxin from the amino end of the alpha chain, generating C5b (beta and alpha' chains).

COMMENT Activation of C5 initiates the spontaneous assembly of the late complement components, C5-C9, into the membrane attack complex. C5b has a transient binding site for C6. The C5b-C6 complex is the foundation upon which the membrane attack complex is assembled.

COMMENT C5a has potent spasmogenic and chemotactic activity.

GENETICS

#map\_position 2

#introns 22/3; 86/3; 140/3; 164/3; 195/2; 223/1; 253/2; 291/3; 334/1; 372/3; 434/3; 502/3; 572/3; 622/3; 667/1; 691/1; 757/1; 787/2; 812/1; 858/3; 934/3; 955/1; 1056/1; 1081/2; 1134/3; 1166/3; 1224/1; 1292/3; 1343/3; 1364/3; 1392/1; 1411/2; 1445/3; 1470/3; 1506/1; 1534/1; 1564/1; 1592/1; 1637/2

CLASSIFICATION #superfamily alpha-2-macroglobulin

KEYWORDS complement alternate pathway; complement pathway; cytolysis; glycoprotein; inflammation; membrane attack complex; plasma

FEATURE

1-18 #domain signal sequence #status predicted #label SIG\

19-674,679-1679 #product complement C5 #status predicted #label MAT\

19-674,756-1679 #product C5b #status predicted #label C5B\

19-674 #product complement C5 and C5b beta chain #status predicted #label C5BB\

679-1679 #product complement C5 alpha chain #status predicted #label C5A\

679-755 #product C5a anaphylatoxin #status predicted #label C5T\

756-1679 #product C5b alpha' chain #status predicted #label C5BA\

567-814,635-670,702-728,703-735,715-736,870-1531,1105-1163,1379-1509,1409-1478,1524-1529,1536-1609,1557-1679,1657-1660

915.1119,1633 #disulfide\_bonds #status predicted\

#binding\_site carbohydrate (Asn) (covalent) #status predicted

SUMMARY #length 1680 #molecular-weight 188876 #checksum 3888

Query Match 71.4%; Score 50; DB 1; Length 1680;

Best Local Similarity 62.5%; Pred. No. 1.43e+01;

Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 833 IPYSVVRG 840

QY 1 IPYPVIRS 8

```

#cross-references EMBL:X54309
#experimental_source strain K-12
#note this sequence has been proven to be erroneous in
Ref:S27311

REFERENCE
A40661
McDowall, K.J.; Hernandez, R.G.; Lin-Chao, S.; Cohen, S.N.
J. Bacteriol. (1993) 175:4245-4249
The ams-1 and rne-3071 temperature-sensitive mutations in the
ams gene are in close proximity to each other and cause
substitutions within a domain that resembles a product of
the Escherichia coli rne locus.
#cross-references MUID:93308106
#accession A40661
#status not compared with conceptual translation
#molecule_type DNA
#residues 1-486, 'V', 488-489 #label MCD
#note sequence extracted from NCBI backbone (NCBIP:134520)
GENETICS
#gene rne; ams; hmp1
#map_position 24 min
#description cleaves RNA endonucleolytically in AU-rich single-strand
regions; RNA-binding activity; interacts with
polynucleotide phosphorylase and other proteins implicated
in processing and degradation of RNA
#note autoregulation
CLASSIFICATION
#superfamily ribonuclease E
KEYWORDS
endonuclease; hydrolase; P-loop; phosphoric diester
hydrolase; RNA binding; transmembrane protein
FEATURE
113-131 #domain transmembrane #status predicted #label TM\
169-176 #region nucleotide-binding motif A (P-loop) #status
atypical\
524-568 #region proline-rich\
743-778 #region proline-rich\
SUMMARY
#length 1061 #molecular-weight 118196 #checksum 5236
Query Match 71.4%; Score 50; DB 1; Length 1061;
Best Local Similarity 75.0%; Pred. No. 1.43e+01;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 844 TRYPIVRP 851
| | | | |
| | | | |
Qy 1 IPYPIVRS 8

RESULT 11
ENTRY T01334 #type complete
TITLE hypothetical protein F6N15.10 - Arabidopsis thaliana
ORGANISM #formal_name Arabidopsis thaliana #common_name mouse-ear
cress
DATE 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change
12-Feb-1999
ACCESSIONS T01334
REFERENCE 214297
#authors Ryan, E.; Edwards, J.; Pape, K.
#submission submitted to the EMBL Data Library, May 1998
#description The sequence of A. thaliana F6N15.
#accession T01334
#status preliminary; translated from GB/EMBL/DDBJ
#molecule_type DNA
#residues 1-1260 #label RYA
#cross-references EMBL:AF069299; NID:g3193311; PID:g3193327
GENETICS
#map_position IV
#introns 218/2; 303/3; 340/3; 935/3; 962/1; 1061/1; 1128/3; 1212/3
#note F6N15.10
SUMMARY
#length 1260 #molecular-weight 139283 #checksum 2898
Query Match 71.4%; Score 50; DB 2; Length 1260;
Best Local Similarity 55.6%; Pred. No. 1.43e+01;
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

```

```

Db 784 LPYPILRPI 792
| | | | |
| | | | |
Qy 1 IPYPIVRS 9

RESULT 12
ENTRY C5HU #type complete
TITLE Complement C5 precursor - human
CONTAINS C5a anaphylatoxin; C5b
ORGANISM #formal_name Homo sapiens #common_name man
DATE 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change
24-Oct-1997
ACCESSIONS A40075; A27689; A01267; A01266; S15121
REFERENCE A40075
#authors Haviland, D.L.; Haviland, J.C.; Fleischer, D.T.; Hunt, A.;
Wetsel, R.A.
#journal J. Immunol. (1991) 146:362-368
#title Complete cDNA sequence of human complement pro-C5. Evidence
of truncated transcripts derived from a single copy gene.
#cross-references MUID:91079575
#accession A40075
#molecule_type mRNA
#residues 1-1676 #label HAV
#cross-references GB:M57725; NID:g179982; PID:g179983
#note 518-Ser was also found
REFERENCE A27689
#authors Wetsel, R.A.; Lemons, R.S.; Le Beau, M.M.; Barnum, S.R.;
Noack, D.; Tack, B.F.
#journal Biochemistry (1988) 27:1474-1482
#title Molecular analysis of human complement component C5:
localization of the structural gene to chromosome 9.
#cross-references MUID:88209511
#accession A27689
#molecule_type mRNA
#residues 412-1676 #label WET
#cross-references GB:M65134; GB:M18879; NID:g179691; PID:g179692
REFERENCE A01267
#authors Fernandez, H.N.; Hugli, T.E.
#journal J. Biol. Chem. (1978) 253:6955-6964
#title Primary structural analysis of the polypeptide portion of
human C5a anaphylatoxin. Polypeptide sequence determination
and assignment of the oligosaccharide attachment site in
C5a.
#cross-references MUID:79005687
#accession A01267
#molecule_type protein
#residues 678-751 #label FER
REFERENCE A01266
#authors Lundwall, A.B.; Wetsel, R.A.; Kristensen, T.; Whitehead,
A.S.; Woods, D.E.; Ogden, R.C.; Colten, H.R.; Tack, B.F.
#journal J. Biol. Chem. (1985) 260:2108-2112
#title Isolation and sequence analysis of a cDNA clone encoding the
fifth complement component.
#cross-references MUID:85130937
#accession A01266
#molecule_type mRNA
#residues 412-854,
'SIALSPRLCNGKISGHCKLRPLPGSSDPSASQVAGITGTHHHAQPT',
#label LUN
#cross-references GB:K02874
#note the carboxyl-terminal part of the sequence in this
report appears to be derived from translation of an
ALU repeat sequence
REFERENCE S15121
#authors Bohnsack, J.F.; Mollison, K.W.; Buko, A.M.; Ashworth, J.C.;
Hill, H.R.
#journal Biochem. J. (1991) 273:635-640
#title Group B streptococci inactivate complement component C5a by
enzymic cleavage at the C-terminus.
#cross-references MUID:91144547
#contents annotation
#comment Complement C5 contains two disulfide-linked chains, formed by
removal of four basic residues. C5 convertase releases C5a

```

```

TITLE      recombination protein RecN - Aquifex aeolicus
ORGANISM   #formal_name Aquifex aeolicus
DATE       08-May-1998 #sequence_revision 08-May-1998 #text_change
ACCESSIONS F70350
REFERENCE   #authors
#journal   Lenox, A.L.; Graham, D.E.; Overbeek, R.; Shead, M.A.;
#title     Keller, M.; AuJay, M.; Huber, R.; Feldman, R.A.; Short,
           J.M.; Olson, G.J.; Swanson, R.V.
           Nature (1998) 392:353-358
           The complete genome of the hyperthermophilic bacterium
           Aquifex aeolicus.
#cross-references MUID:9819666
#accession F70350
##status   preliminary; nucleic acid sequence not shown;
           translation not shown
##molecule_type DNA
##residues 1-520 ##label AOF
##cross-references GB:AE000695; NID:g2983180; PID:g2983189; GB:AE000657
##experimental_source strain VF5
GENETICS
#gene      recN
#summary   #length 520 #molecular-weight 60439 #checksum 5355
Query Match 72.9%; Score 51; DB 2; Length 520;
Best Local Similarity 66.7%; Pred. NO. 9.21e+00;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
Db 480 IPYIVREL 488
Qy 1 IPYPIVRL 9
RESULT 9
ENTRY
TITLE     T02532 #type complete
#journal   hypothetical protein F13M22.16 - Arabidopsis thaliana
#title     #formal_name Arabidopsis thaliana #common_name mouse-ear
           cross
DATE       05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change
ACCESSIONS T02532
REFERENCE   #authors
#journal   Rounsley, S.D.; Kaul, S.; Lin, X.; Ketchum, K.A.; Crosby,
#title     M.L.; Brandon, R.C.; Sykes, S.M.; Mason, T.M.; Kerlavage,
           A.R.; Adams, M.D.; Somerville, C.R.; Venter, J.C.
           submitted to the EMBL Data Library, June 1998
#description Arabidopsis thaliana chromosome II BAC F13M22 genomic
           sequence.
#accession T02532
##status   preliminary; translated from GB/EMBL/DBDJ
##molecule_type DNA
##residues 1-337 ##label ROU
##cross-references EMBL:AC004684; NID:g3236234; PID:g3236248
GENETICS
#map_position 2
#introns      89/1; 179/3; 192/1; 236/3; 257/2; 298/3
#note         F13M22.16
#summary      #length 337 #molecular-weight 36085 #checksum 5003
Query Match 71.4%; Score 50; DB 2; Length 337;
Best Local Similarity 62.5%; Pred. NO. 1.43e+01;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Db 251 IPYITIRA 258
Qy 1 IPYPIVRS 8
RESULT 10
ENTRY
TITLE     S27311 #type complete
#journal   ribonuclease E (EC 3.1.4.-) - Escherichia coli
#title     cell shape-determining protein; message stability-altering
           ALTERNATE_NAMES

```

```

protein; RNase E
#formal_name Escherichia coli
#sequence_revision 05-Dec-1997 #text_change
DATE       26-Feb-1999
ACCESSIONS A64852; S45572; S27311; A23747; JG0009; A40661; S13127;
           S25116
REFERENCE   #authors
#journal   Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.;
#title     Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.;
           Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.;
           Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao,
           Y.
           Science (1997) 277:1453-1462
           The complete genome sequence of Escherichia coli K-12.
#cross-references MUID:97426617
#accession A64852
##status   nucleic acid sequence not shown; translation not shown
##molecule_type DNA
##residues 1-1061 ##label BLAT
##cross-references GB:AE000209; GB:U00096; NID:gl787322; PID:gl787325;
           UWGP:B1084
##experimental_source strain K-12, substrain MG1655
REFERENCE   #authors
#journal   Casaregola, S.; Jacq, A.; Laoudj, D.; McGurk, G.; Margaron,
#title     S.; Tempete, M.; Norris, V.; Holland, I.B.
           J. Mol. Biol. (1994) 238:867
#cross-references MUID:94238701
#accession S45572
##molecule_type DNA
##residues 1001-1061 ##label CAS
REFERENCE   #authors
#journal   Casaregola, S.; Jacq, A.; Laoudj, D.; McGurk, G.; Margaron,
#title     S.; Tempete, M.; Norris, V.; Holland, I.B.
           J. Mol. Biol. (1992) 238:30-40
#cross-references MUID:94238701
#accession S45572
##molecule_type DNA
##residues 1001-1061 ##label CAS
#accession S27311
##molecule_type DNA
##residues 1-486, 'V', 488-563, 'R', 565-783, 'K', 785-904, 'R', 906-1000,
           1060-1061, 'ITTLPANDARSTGICSGATASQ', ##label CA2
##cross-references EMBL:X67470; NID:949115; PID:g49116
##experimental_source strain MC4100
REFERENCE   #authors
#journal   Claverie-Martin, F.; Diaz-Torres, M.R.; Yancey, S.D.;
#title     Kushner, S.R.
           J. Biol. Chem. (1991) 266:2843-2851
           Analysis of the altered mRNA stability (ams) gene from
           Escherichia coli. Nucleotide sequence, transcriptional
           analysis, and homology of its product to MRP3, a
           mitochondrial ribosomal protein from Neurospora crassa.
#cross-references MUID:91131576
#accession A23747
##status   preliminary
##molecule_type DNA
##residues 1-389, 'H', 391-486, 'V', 488-795, 'SF', 798, 1009, 'LASS',
           1014-1015, 'RKWSASSLS', ##label CLA
##cross-references GB:M36288; GB:M62747; NID:g145271; PID:g145273
##experimental_source strain K-12
##note      this sequence has been proven to be erroneous in
           Ref:S27311
REFERENCE   #authors
#journal   Chauhan, A.K.; Miczak, A.; Taraseviciene, L.; Apirion, D.
#title     Nucleic Acids Res. (1991) 19:125-129
           Sequencing and expression of the rne gene of Escherichia
           coli.
#cross-references MUID:91187608
#accession JG0009
##status   preliminary
##molecule_type DNA
##residues 1-258, 'N', 260-529, 'QPLPCR', 'MC', 719, 'LR', 722-726, 'LPRLL',
           ##label CHA

```

Hornsby, T.; Jagels, K.; Krogh, A.; McLean, J.; Moule, S.;  
Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.;  
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.;  
Skelton, S.; Squares, S.; Squires, R.; Sulston, J.E.;  
Taylor, K.; Whitehead, S.; Barrell, B.G.  
#journal Nature (1998) 393:537-544  
#title Deciphering the biology of Mycobacterium tuberculosis from  
the complete genome sequence.  
#cross-references MUID:98295987  
#accession A70520  
##status preliminary; nucleic acid sequence not shown;  
translation not shown  
##molecule\_type DNA  
##residues 1-398 #label COL  
##cross-references GB:297193; GB:AL123456; NID:93261816; PID:e324824;  
PID:g2225944  
##experimental\_source strain H37Rv  
GENETICS  
#gene aceAb  
SUMMARY  
#length 398 #molecular-weight 44581 #checksum 3240  
Query Match 72.9%; Score 51; DB 2; Length 398;  
Best Local Similarity 66.7%; Pred. No. 9.21e+00;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
Db 71 IPYAIKSL 79  
QY 1 IPYIVRSL 9  
6  
RESULT ENTRY  
TITLE Probable Acyl-CoA Dehydrogenas - Mycobacterium tuberculosis  
(strain H37Rv)  
ORGANISM #formal\_name Mycobacterium tuberculosis  
DATE 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change  
17-Jul-1998  
ACCESSIONS D70884  
REFERENCE A70500  
#authors Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher,  
C.; Harris, D.; Gordon, S.V.; Eigmeier, K.; Gas, S.; Barry  
III, C.E.; Tekala, F.; Badcock, K.; Basham, D.; Brown, D.;  
Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.;  
Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;  
Hornsby, T.; Jagels, K.; Krogh, A.; McLean, J.; Moule, S.;  
Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.;  
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.;  
Skelton, S.; Squares, S.; Squires, R.; Sulston, J.E.;  
Taylor, K.; Whitehead, S.; Barrell, B.G.  
#journal Nature (1998) 393:537-544  
#title Deciphering the biology of Mycobacterium tuberculosis from  
the complete genome sequence.  
#cross-references MUID:98295987  
#accession D70884  
##status preliminary; nucleic acid sequence not shown;  
translation not shown  
##molecule\_type DNA  
##residues 1-410 #label COL  
##cross-references GB:AL008967; GB:AL123456; NID:93261491; PID:el173919;  
PID:g2624311  
##experimental\_source strain H37Rv  
GENETICS  
#gene fadE21  
SUMMARY  
#length 410 #molecular-weight 44743 #checksum 7911  
Query Match 72.9%; Score 51; DB 2; Length 410;  
Best Local Similarity 75.0%; Pred. No. 9.21e+00;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
Db 37 PYIARKL 44  
QY 2 PYIVRSL 9

7  
RESULT

ENTRY

TITLE

ORGANISM

DATE

ACCESSIONS

REFERENCE

#authors

F69811 #type complete

2-oxoglutarate/malate translocator homolog yfls - Bacillus

subtilis

#formal\_name Bacillus subtilis

05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change

F69811

24-Sep-1998

A69580

Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.;

Alloni, G.; Azevedo, V.; Bertero, M.G.; Bessieres, P.;

Bolotin, A.; Borchert, S.; Boriss, R.; Boursier, L.; Brans,

A.; Braun, M.; Brignell, S.C.; Bron, S.; Brouillet, S.;

Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;

Choi, S.K.; Codani, J.J.; Connerton, I.F.; Cummings, N.J.;

Daniel, R.A.; Denizot, F.; Devine, K.M.; Duesterhoeft, A.;

Enflich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.;

Fabret, C.; Ferrari, E.; Foulger, D.; Fritz, C.; Fujita,

M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galleron, N.; Ghim,

S.Y.; Glaser, P.; Goffeau, A.; Golightly, E.J.; Grandi, G.;

Guiseppi, G.; Guy, B.J.; Haga, K.; Halech, J.; Harwood,

C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.;

Hullo, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karamata, D.;

Kasahara, Y.; Klaerr-Blanchard, M.; Klein, C.; Kobayashi,

Y.; Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.;

Kurita, K.; Lapidus, A.; Lardinois, S.; Lauber, J.;

Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.;

Maueel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno,

M.; Moestl, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly,

M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro,

V.; Pohl, T.M.; Portetelle, D.; Porwollik, S.; Prescott,

A.M.; Prescan, E.; Pujic, P.; Purnelle, B.; Rapoport, G.;

Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.;

Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon, E.;

Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.;

Sekowska, A.; Seror, S.J.; Serror, P.; Shin, B.S.; Soldo,

B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Takahashi, H.;

Takemaru, K.; Takeuchi, M.; Tamakoshi, A.; Tanaka, T.;

Terpsira, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.;

Vandenbol, M.; Vannier, F.; Vassarotti, A.; Viari, A.;

Wambutt, R.; Wedler, E.; Wedler, H.; Weitzenecker, T.;

Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto,

K.; Yata, K.; Yoshida, K.; Yoshikawa, H.F.; Zumstein, E.;

Yoshikawa, H.; Danchin, A.

Nature (1997) 390:249-256

The complete genome sequence of the Gram-positive bacterium

Bacillus subtilis.

#cross-references MUID:98044033

#accession F69811

#status preliminary; nucleic acid sequence not shown;

translation not shown

#molecule\_type DNA

#residues 1-478 #label KUN

#cross-references GB:299108; GB:AL009126; NID:92633055; PID:el182747;

PID:g2633081

#experimental\_source strain 168

GENETICS

#gene yfls

CLASSIFICATION

#superfamily 2-oxoglutarate/malate translocator

SUMMARY

#length 478 #molecular-weight 51431 #checksum 768

Query Match 72.9%; Score 51; DB 2; Length 478;

Best Local Similarity 66.7%; Pred. No. 9.21e+00;

Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 155 IIFPIRSL 163

QY 1 IPYIVRSL 9

8  
RESULT ENTRY

F70350 #type complete

```

##molecule_type mRNA
##residues 1-209 #label ROB
##note based on the evidence for Gln-tRNA, the authors translated the codon TAG as Gln; the sequence shown follows the authors' translation
CLASSIFICATION #superfamily ras transforming protein; translation elongation factor Tu homology
KEYWORDS GTP binding; P-loop
FEATURE
10-17 #region nucleotide-binding motif A (P-loop)\
140-143 #region GTP-binding NKXD motif\
168-170 #region GTP-binding SAK/L motif\
16,17,58,140,141, #binding_site Mg-GTP (Lys, Ser, Thr, Asn, Lys, Asp, Ser)
143,168 #status predicted
SUMMARY #length 209 #molecular-weight 23854 #checksum 3860
Query Match 75.7%; Score 53; DB 2; Length 209;
Best Local Similarity 66.7%; Pred. No. 3.74e+00;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Db 178 IPXSLVREL 186
QY 1 IPYPIVRSLS 9
RESULT 3
ENTRY QOECRS #type complete
TITLE YgJE protein - Escherichia coli
ORGANISM #formal_name Escherichia coli
DATE 30-Jun-1998 #sequence_revision 31-Oct-1997 #text_change 17-Jul-1998
ACCESSION E65094; C29049
REFERENCE A64720
#authors Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.; Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.; Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao, Y.
#journal Science (1997) 277:1453-1462
#title The complete genome sequence of Escherichia coli K-12.
#cross-references MUID:97426617
#accession E65094
##status nucleic acid sequence not shown; translation not shown
##molecule_type DNA
##residues 1-487 #label BLAT
##cross-references GB:AE000388; GB:U00096; NID:g1789441; PID:g1789444; UWGP:b3063
##experimental_source strain K-12, substrain MG1655
REFERENCE A91573
#authors Nesin, M.; Lupski, J.R.; Svec, P.; Godson, G.N.
#journal Gene (1987) 51:149-161
#title Possible new genes as revealed by molecular analysis of a 5-kb Escherichia coli chromosomal region 5' to the rpsU-dnaG-rpoD macromolecular-synthesis operon.
#cross-references MUID:87248073
#accession C29049
##molecule_type DNA
##residues 279-403, 'P', 405-411, 'RWRCKSRKRCSEA' #label NES
GENETICS
#gene ygJE
#map_position 67 min
CLASSIFICATION #superfamily 2-oxoglutarate/malate translocator
KEYWORDS transmembrane protein
FEATURE
11-27 #domain transmembrane #status predicted #label TM1\
33-49 #domain transmembrane #status predicted #label TM2\
52-68 #domain transmembrane #status predicted #label TM3\
95-111 #domain transmembrane #status predicted #label TM4\
138-154 #domain transmembrane #status predicted #label TM5\
206-222 #domain transmembrane #status predicted #label TM6\
237-253 #domain transmembrane #status predicted #label TM7\
289-305 #domain transmembrane #status predicted #label TM8\

```

```

310-326 #domain transmembrane #status predicted #label TM9\
378-394 #domain transmembrane #status predicted #label TM10\
422-438 #domain transmembrane #status predicted #label TM11\
464-480 #domain transmembrane #status predicted #label TM12\
SUMMARY #length 487 #molecular-weight 52906 #checksum 1643
Query Match 75.7%; Score 53; DB 1; Length 487;
Best Local Similarity 66.7%; Pred. No. 3.74e+00;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Db 165 IYPIIRNL 173
QY 1 IPYPIVRSLS 9
RESULT 4
ENTRY C64416 #type complete
TITLE conserved hypothetical MG372 related protein - Methanococcus jannaschii
ORGANISM #formal_name Methanococcus jannaschii
DATE 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 18-Sep-1998
ACCESSION C64416
REFERENCE A64300
#authors Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake, J.A.; Fitzgerald, L.M.; Clayton, R.A.; Gocayne, J.D.; Kerlavage, A.R.; Dougherty, B.A.; Tomb, J.F.; Adams, M.D.; Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.; Scott, J.L.; Geoghegan, N.S.M.; Weidman, J.F.; Fuhrman, J.L.; Nguyen, D.; Uterback, T.R.; Kelley, J.M.; Peterson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.; Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese, C.R.; Venter, J.C.
#journal Science (1996) 273:1058-1073
#title Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii.
#cross-references MUID:96337999
#accession C64416
##status preliminary; nucleic acid sequence not shown; translation not shown
##molecule_type DNA
##residues 1-381 #label BUL
##cross-references GB:U67536; GB:L77117; NID:gl591596; PID:gl591602; TIGR:MJ0931; PID:gl510973
GENETICS
#map_position FOR860923-862068
CLASSIFICATION #superfamily Mycoplasma genitalium hypothetical protein MG372
SUMMARY #length 381 #molecular-weight 43436 #checksum 7754
Query Match 72.9%; Score 51; DB 2; Length 381;
Best Local Similarity 66.7%; Pred. No. 9.21e+00;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Db 319 INYPIIRPL 327
QY 1 IPYPIVRSLS 9
RESULT 5
ENTRY A70520 #type complete
TITLE probable aceAB protein - Mycobacterium tuberculosis (strain H37RV)
ORGANISM #formal_name Mycobacterium tuberculosis
DATE 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 17-Jul-1998
ACCESSION A70520
REFERENCE A70500
#authors Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.V.; Eigmeier, K.; Gas, S.; Barry III, C.E.; Tekala, F.; Badcock, K.; Basham, D.; Brown, D.; Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;

```

\*\*\*\*\*  
[M][P][E][R][L] (TM)  
\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Fri Apr 14 23:49:23 2000; MasPar time 3.32 Seconds  
Tabular output not generated. 108.704 Million cell updates/sec

Title: >US-08-452-843-9  
Description: (1-9) from US08452843.ppep  
Perfect Score: 70  
Sequence: 1 IPYPIVRSLS 9

Scoring table: PAM 150  
Gap 15

Searched: 122810 seqs, 40068593 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: pir62  
1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 24.042; Variance 30.857; scale 0.779

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES				
Result No.	Score	Query Match	Length DB ID	Description
1	56	80.0	340 2	S62493 hypothetical protein
2	53	75.7	209 2	S13179 transforming protein
3	53	75.7	487 1	Q9ECSR ygjE protein - Escher
4	51	72.9	381 2	C64416 conserved hypothetical
5	51	72.9	398 2	A70520 probable aceAb protei
6	51	72.9	410 2	D70884 probable Acyl-CoA Deh
7	51	72.9	478 2	F69811 2-oxoglutarate/mala
8	51	72.9	520 2	F70350 recombination protein
9	50	71.4	337 2	T02532 hypothetical protein
10	50	71.4	1061 1	S27311 ribonuclease E (ec 3.
11	50	71.4	1260 2	T01334 hypothetical protein
12	50	71.4	1676 1	C5HU complement C5 precurs
13	50	71.4	1680 1	C5MS complement C5 precurs
14	50	71.4	4344 1	A53489 dynein heavy chain, c
15	50	71.4	4367 2	B54802 dynein heavy chain, c
16	49	70.0	57 2	D35826 hypothetical protein
17	49	70.0	219 2	S75541 hypothetical protein
18	49	70.0	238 2	B64313 probable 3-isopropylm
19	49	70.0	298 1	MMAGCF membrane protein lacP
20	49	70.0	567 2	A71463 probable sulfate tran
21	48	68.6	78 2	C64472 hypothetical protein
22	48	68.6	305 2	S35991 C-alpha-dehydrogenase
23	48	68.6	432 2	G69993 conserved hypothetical

ALIGNMENTS

RESULT 1  
ENTRY S62493 #type complete  
TITLE hypothetical protein SPAC23D3.02 - fission yeast  
ORGANISM (Schizosaccharomyces pombe)  
DATE 16-May-1996 #sequence\_revision 13-Mar-1997 #text\_change 21-Aug-1998

ACCESSIONS S62493  
REFERENCE S62492  
#authors Niblett, D.; Harris, D.  
#submission submitted to the EMBL Data Library, October 1995  
#accession S62493  
#status preliminary  
#molecule\_type DNA  
#residues 1-340 #label NIB  
#cross-references EMBL:Z64354; NID:g1039338; PID:g1039340

GENETICS  
#map\_position 1R  
#introns 25/2

CLASSIFICATION #superfamily phage T4 DNA polymerase accessory protein 44  
SUMMARY #length 340 #molecular-weight 37876 #checksum 2036

Query Match 80.0%; Score 56; DB 2; Length 340;  
Best Local Similarity 66.7%; Pred. No. 9.25e+01;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 249 VPYNIIRSL 257  
QY 1 IPYPIVRSLS 9  
:||| |::|||

RESULT 2  
ENTRY S13179 #type complete  
TITLE transforming protein (ras) - Geodia cydonium  
ORGANISM #formal\_name Geodia cydonium  
DATE 19-Mar-1997 #sequence\_revision 19-Mar-1997 #text\_change 19-Dec-1998

ACCESSIONS S13179  
REFERENCE S13179  
#authors Robitzki, A.; Schroeder, H.C.; Ugarkovic, D.; Kuchino, Y.; Kurelec, B.; Gamulin, V.; Mueller, W.E.G.  
#journal Eur. J. Biochem. (1990) 192:499-506  
#title Regulated expression and phosphorylation of the 23-26-kDa ras protein in the sponge Geodia cydonium.  
#cross-references MUID:91006138  
#accession S13179  
#status preliminary

\*\*\*\*\*  
WILEY  
\*\*\*\*\*  
(TM)  
\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Sat Apr 15 00:02:53 2000; MasPar time 13.78 Seconds  
45.280 Million cell updates/sec

Tabular output not generated.

Title: >US-08-452-843-11  
Description: (1-9) from US08452843.pep  
Perfect Score: 69  
Sequence: 1 RYRPGTVAL 9

Scoring table: PAM 150  
Gap 15

Searched: 225978 seqs, 69334122 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: sptrembl12  
1:sp\_archaea 2:sp\_bacteria 3:sp\_fungi 4:sp\_human  
5:sp\_invertebrate 6:sp\_mammal 7:sp\_mhc 8:sp\_organelle  
9:sp\_phage 10:sp\_plant 11:sp\_rodent 12:sp\_unclassified  
13:sp\_vertebrate 14:sp\_virus

Statistics: Mean 23.560; Variance 25.040; scale 0.941

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES				Pred. No.	
Result No.	Score	Match	Description	ID	Pred. No.
1	69	100.0	HISTONE H3 (FRAGMENT).	Q42775	5.72e-05
2	69	100.0	HISTONE H3 (FRAGMENT).	Q42787	5.72e-05
3	69	100.0	HISTONE H3 (FRAGMENT).	Q43313	5.72e-05
4	69	100.0	HISTONE H3 (FRAGMENT).	Q43031	5.72e-05
5	69	100.0	HISTONE H3 (FRAGMENT).	Q42723	5.72e-05
6	69	100.0	HISTONE H3 (FRAGMENT).	Q3XHM9	5.72e-05
7	69	100.0	HISTONE H3 (FRAGMENT).	Q41612	5.72e-05
8	69	100.0	HISTONE H3 (FRAGMENT).	Q42736	5.72e-05
9	69	100.0	HISTONE H3 (FRAGMENT).	Q42745	5.72e-05
10	69	100.0	HISTONE H3-D (FRAGMENT).	Q3XHN8	5.72e-05
11	69	100.0	HISTONE H3-D (FRAGMENT).	Q3XHN2	5.72e-05
12	69	100.0	HISTONE H3-D (FRAGMENT).	Q3XHN7	5.72e-05
13	69	100.0	HISTONE H3 (FRAGMENT).	Q42826	5.72e-05
14	69	100.0	HISTONE H3 (FRAGMENT).	Q27899	5.72e-05
15	69	100.0	HISTONE H3 (FRAGMENT).	P91947	5.72e-05
16	69	100.0	HISTONE H3 (FRAGMENT).	O61402	5.72e-05
17	69	100.0	HISTONE H3 (FRAGMENT).	O61405	5.72e-05
18	69	100.0	HISTONE H3 (FRAGMENT).	O61403	5.72e-05
19	69	100.0	HISTONE H3 (FRAGMENT).	Q3XV15	5.72e-05
20	69	100.0	HISTONE H3 (FRAGMENT).	Q29569	5.72e-05

21	69	100.0	5	Q9XYK2	HISTONE H3 (FRAGMENT).	5.72e-05
22	69	100.0	105	Q9XYI6	HISTONE H3 (FRAGMENT).	5.72e-05
23	69	100.0	106	Q9XYI0	HISTONE H3 (FRAGMENT).	5.72e-05
24	69	100.0	106	Q9XYI7	HISTONE H3 (FRAGMENT).	5.72e-05
25	69	100.0	107	Q9XYJ6	HISTONE H3 (FRAGMENT).	5.72e-05
26	69	100.0	107	Q9XYI1	HISTONE H3 (FRAGMENT).	5.72e-05
27	69	100.0	108	Q9XYU8	HISTONE H3 (FRAGMENT).	5.72e-05
28	69	100.0	108	Q9XYJ3	HISTONE H3 (FRAGMENT).	5.72e-05
29	69	100.0	108	Q9XYI8	HISTONE H3 (FRAGMENT).	5.72e-05
30	69	100.0	108	Q9XYK4	HISTONE H3 (FRAGMENT).	5.72e-05
31	69	100.0	109	Q9XYJ5	HISTONE H3 (FRAGMENT).	5.72e-05
32	69	100.0	109	Q9XYJ9	HISTONE H3 (FRAGMENT).	5.72e-05
33	69	100.0	109	Q9XYI2	HISTONE H3 (FRAGMENT).	5.72e-05
34	69	100.0	109	Q9XYJ4	HISTONE H3 (FRAGMENT).	5.72e-05
35	69	100.0	109	Q9XYJ7	HISTONE H3 (FRAGMENT).	5.72e-05
36	69	100.0	109	Q9XYK0	HISTONE H3 (FRAGMENT).	5.72e-05
37	69	100.0	115	P90676	HISTONE H3 (FRAGMENT).	5.72e-05
38	69	100.0	127	Q42832	HISTONE H3 (FRAGMENT).	5.72e-05
39	69	100.0	136	Q92068	HISTONE H3.	5.72e-05
40	69	100.0	136	O64528	HISTONE H3.	5.72e-05
41	69	100.0	136	Q27718	HISTONE H3.	5.72e-05
42	69	100.0	136	O02648	HISTONE H3.	5.72e-05
43	69	100.0	136	Q27866	HISTONE H3.	5.72e-05
44	69	100.0	136	Q27719	HISTONE H3.	5.72e-05
45	69	100.0	136	Q92133	HISTONE H3.	5.72e-05

ALIGNMENTS

RESULT 1  
ID Q42775 PRELIMINARY; PRT; 56 AA.  
AC Q42775; Q42771; Q42774;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)  
DE HISTONE H3 (FRAGMENT).  
GN HIS3.  
OS Glycine latifolia.  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;  
OC core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae;  
OC Glycine.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 97131650.  
RA DOYLE J.J., KANAZIN V., SHOEMAKER R.C.;  
RT "Phylogenetic utility of histone H3 intron sequences in the perennial relatives of soybean (Glycine: Leguminosae).";  
RL Mol. Phylogenet. Evol. 6:438-447(1996).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV. G2545, CV. G1137;  
RX MEDLINE; 99261647.  
RA DOYLE J.J., DOYLE J.L., BROWN A.H.;  
RT "Incongruence in the diploid B-genome species complex of Glycine (Leguminosae) revisited: histone H3-D alleles versus chloroplast haplotypes.";  
RL Mol. Biol. Evol. 16:354-362(1999).  
DR EMBL; U47404; AAB50475.1; -;  
DR EMBL; U47368; AAB50457.1; -;  
DR EMBL; U47387; AAB50475.1; -;  
DR EMBL; AF093435; AAD40987.1; -;  
DR EMBL; AF093434; AAD40986.1; -;  
DR MENDEL; 30803; Glyla:His3:30803.  
KW Nuclear protein; Chromosomal protein; DNA-binding; Nucleosome core.  
FT NON\_TER  
SQ SEQUENCE 56 AA; 6506 MW; AA28B116 CRC32;

Query Match 100.0%; Score 69; DB 10; Length 56;  
Best Local Similarity 100.0%; Pred.No. 5.72e-05;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 17 RYRPGTVAL 25  
QY 1 RYRPGTVAL 9

RESULT 2  
ID Q42787 PRELIMINARY; PRT; 56 AA.  
AC Q42787; Q42791;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)  
DE HISTONE H3 (FRAGMENT).  
GN HIS3.  
OS Glycine microphylla.  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC euryliophytes; Spermatophyta; Magnoliophyta; eudicotyledons;  
OC core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae;  
OC Glycine.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 97131650.  
RA DOYLE J.J., KANAZIN V., SHOEMAKER R.C.;  
RT "Phylogenetic utility of histone H3 intron sequences in the perennial  
relatives of soybean (Glycine: Leguminosae).";  
RL Mol. Phylogenet. Evol. 6:438-447(1996).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-CV. G1498. CV. G1143;  
RX MEDLINE: 99261647.  
RA DOYLE J.J., DOYLE J.L., BROWN A.H.;  
RT "Incongruence in the diploid B-genome species complex of Glycine  
(Leguminosae) revisited: histone H3-D alleles versus chloroplast  
haplotypes.";  
RL Mol. Biol. Evol. 16:354-362(1999).  
CC -!- FUNCTION: HISTONE H3, ALONG WITH HISTONE H4, PLAYS A CENTRAL ROLE  
IN NUCLEOSOME FORMATION.  
CC -!- SUBUNIT: THE NUCLEOSOME IS AN OCTAMER CONTAINING TWO MOLECULES OF  
H2A, H2B, H3, AND H4; WHICH WRAP APPROXIMATIVELY 146 BP OF DNA.  
DR EMBL: U47372; AAB50461.1; -;  
DR EMBL: U47407; AAB50488.1; -;  
DR EMBL: AF093438; AAD40990.1; -;  
DR EMBL: AF093436; AAD40988.1; -;  
DR MENDEL; 15669; Glymi;HIS3;15669.  
KW Nuclear protein; Chromosomal protein; DNA-binding; Nucleosome core.  
FT NON\_TER 1 56  
FT NON\_TER 56  
SQ SEQUENCE 56 AA; 6506 MW; AA28B116 CRC32;

Query Match 100.0%; Score 69; DB 10; Length 56;  
Best Local Similarity 100.0%; Pred. No. 5.72e-05;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 17 RYRPGTVAL 25  
QY 1 RYRPGTVAL 9

RESULT 3  
ID Q43313 PRELIMINARY; PRT; 56 AA.  
AC Q43313;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)  
DE HISTONE H3 (FRAGMENT).  
GN HIS3.  
OS Glycine tomentella;  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC euryliophytes; Spermatophyta; Magnoliophyta; eudicotyledons;  
OC core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae;  
OC Glycine.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 97131650.  
RA DOYLE J.J., KANAZIN V., SHOEMAKER R.C.;

RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: HISTONE H3, ALONG WITH HISTONE H4, PLAYS A CENTRAL ROLE  
IN NUCLEOSOME FORMATION.  
CC -!- SUBUNIT: THE NUCLEOSOME IS AN OCTAMER CONTAINING TWO MOLECULES OF  
H2A, H2B, H3, AND H4; WHICH WRAP APPROXIMATIVELY 146 BP OF DNA.  
DR EMBL: U47412; AAB50490.1; -;  
DR EMBL: U47377; AAB50447.1; -;  
DR EMBL: U47395; AAB50484.1; -;  
DR MENDEL; 15682; Glyto;HIS3;15682.  
KW Nuclear protein; Chromosomal protein; DNA-binding; Nucleosome core.  
FT NON\_TER 1 56  
FT NON\_TER 56  
SQ SEQUENCE 56 AA; 6506 MW; AA28B116 CRC32;

Query Match 100.0%; Score 69; DB 10; Length 56;  
Best Local Similarity 100.0%; Pred. No. 5.72e-05;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 17 RYRPGTVAL 25  
QY 1 RYRPGTVAL 9

RESULT 4  
ID Q43031 PRELIMINARY; PRT; 56 AA.  
AC Q43031;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)  
DE HISTONE H3 (FRAGMENT).  
GN HIS3.  
OS Pseudemania comosa.  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC euryliophytes; Spermatophyta; Magnoliophyta; eudicotyledons;  
OC core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae;  
OC Pseudemania.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA DOYLE J.J., KANAZIN V., SHOEMAKER R.C.;  
RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: HISTONE H3, ALONG WITH HISTONE H4, PLAYS A CENTRAL ROLE  
IN NUCLEOSOME FORMATION.  
CC -!- SUBUNIT: THE NUCLEOSOME IS AN OCTAMER CONTAINING TWO MOLECULES OF  
H2A, H2B, H3, AND H4; WHICH WRAP APPROXIMATIVELY 146 BP OF DNA.  
DR EMBL: U47408; AAB50491.1; -;  
DR MENDEL; 15680; Pseco;HIS3;15680.  
KW Nuclear protein; Chromosomal protein; DNA-binding; Nucleosome core.  
FT NON\_TER 1 56  
FT NON\_TER 56  
SQ SEQUENCE 56 AA; 6522 MW; AB5A375A CRC32;

Query Match 100.0%; Score 69; DB 10; Length 56;  
Best Local Similarity 100.0%; Pred. No. 5.72e-05;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 17 RYRPGTVAL 25  
QY 1 RYRPGTVAL 9

RESULT 5  
ID Q42723 PRELIMINARY; PRT; 56 AA.  
AC Q42723;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)  
DE HISTONE H3 (FRAGMENT).  
GN HIS3.  
OS Dumasia villosa.  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC euryliophytes; Spermatophyta; Magnoliophyta; eudicotyledons;  
OC core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae;  
OC Dumasia.



```
[1]
RN  SEQUENCE FROM N.A.
RP  MEDLINE; 97131650.
RA  DOYLE J.J., KANAZIN V., SHOEMAKER R.C.;
RT  "Phylogenetic utility of histone H3 intron sequences in the perennial
RL  relatives of soybean (Glycine: Leguminosae).";
RM  Mol. Phylogenet. Evol. 6:438-447(1996).
DR  EMBL; U47363; AAB50449.1; -.
DR  MENDEL; 15641; Dumvi;His3;15641.
KW  Nuclear protein; Chromosomal protein; DNA-binding; Nucleosome core.
FT  NON_TER 1
FT  NON_TER 56
SQ  SEQUENCE 56 AA; 6506 MW; AA28B116 CRC32;

Query Match 100.0%; Score 69; DB 10; Length 56;
Best Local Similarity 100.0%; Pred. No. 5.72e-05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 17 RYRPGTVAL 25
QY 1 RYRPGTVAL 9

RESULT 6
ID Q9XHM9 PRELIMINARY; PRT; 56 AA.
AC Q9XHM9
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DE HISTONE H3-D (FRAGMENT).
OS Glycine sp. G2344.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae;
OC Glycine.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CV. G2344;
RX DOYLE J.J., DOYLE J.L., BROWN A.H.;
RT "Incongruence in the diploid B-genome species complex of Glycine
RM (Leguminosae) revisited: histone H3-D alleles versus chloroplast
RL haplotypes.";
RL Mol. Biol. Evol. 16:354-362(1999).
DR EMBL; AF093447; AAD40999.1; -.
FT NON_TER 1
FT NON_TER 56
SQ SEQUENCE 56 AA; 6506 MW; AA28B116 CRC32;

Query Match 100.0%; Score 69; DB 10; Length 56;
Best Local Similarity 100.0%; Pred. No. 5.72e-05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 17 RYRPGTVAL 25
QY 1 RYRPGTVAL 9

RESULT 7
ID Q41612 PRELIMINARY; PRT; 56 AA.
AC Q41612;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DE HISTONE H3 (FRAGMENT).
GN HIS3.
OS Terminus labialis.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae;
OC Terminus.
RN [1]
RP SEQUENCE FROM N.A.
```

```
RX MEDLINE; 97131650.
RA DOYLE J.J., KANAZIN V., SHOEMAKER R.C.;
RT "Phylogenetic utility of histone H3 intron sequences in the perennial
RL relatives of soybean (Glycine: Leguminosae).";
RM Mol. Phylogenet. Evol. 6:438-447(1996).
DR EMBL; U47394; AAB50486.1; -.
DR MENDEL; 15638; Terla;His3;15638.
FT NON_TER 1
FT NON_TER 56
SQ SEQUENCE 56 AA; 6472 MW; 9077D730 CRC32;

Query Match 100.0%; Score 69; DB 10; Length 56;
Best Local Similarity 100.0%; Pred. No. 5.72e-05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 17 RYRPGTVAL 25
QY 1 RYRPGTVAL 9

RESULT 8
ID Q42756 PRELIMINARY; PRT; 56 AA.
AC Q42756; Q42749; Q42750; Q42752; Q42753;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DE HISTONE H3 (FRAGMENT).
GN HIS3.
OS Glycine cytoloba.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae;
OC Glycine.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 97131650.
RA DOYLE J.J., KANAZIN V., SHOEMAKER R.C.;
RT "Phylogenetic utility of histone H3 intron sequences in the perennial
RL relatives of soybean (Glycine: Leguminosae).";
RM Mol. Phylogenet. Evol. 6:438-447(1996).
DR EMBL; U47400; AAB50473.1; -.
DR EMBL; U47361; AAB50454.1; -.
DR EMBL; U47362; AAB50455.1; -.
DR EMBL; U47381; AAB50440.1; -.
DR EMBL; U47382; AAB50441.1; -.
DR MENDEL; 15659; GLCYV;His3;15659.
KW Nuclear protein; Chromosomal protein; DNA-binding; Nucleosome core.
FT NON_TER 1
FT NON_TER 56
SQ SEQUENCE 56 AA; 6506 MW; AA28B116 CRC32;

Query Match 100.0%; Score 69; DB 10; Length 56;
Best Local Similarity 100.0%; Pred. No. 5.72e-05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 17 RYRPGTVAL 25
QY 1 RYRPGTVAL 9

RESULT 9
ID Q42745 PRELIMINARY; PRT; 56 AA.
AC Q42745; Q42740;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DE HISTONE H3 (FRAGMENT).
GN HIS3.
OS Glycine argyrea.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae;
OC Glycine.
```

```

RN SEQUENCE FROM N.A.
RP MEDLINE: 97131650.
RA DOYLE J.J., KANAZIN V., SHOEMAKER R.C.;
RT "Phylogenetic utility of histone H3 intron sequences in the perennial
RL relatives of soybean (Glycine: Leguminosae).";
RL Mol. Phylogenet. Evol. 6:438-447(1996).
DR EMBL: U47397; AAB50470.1; -.
DR EMBL: U47358; AAB50451.1; -.
DR MENDEL; 30804; Glyar.His3.30804.
KW Nuclear protein; Chromosomal protein; DNA-binding; Nucleosome core.
FT NON_TER 1
FT NON_TER 56
SQ SEQUENCE 56 AA; 6506 MW; AA28B116 CRC32;

Query Match 100.0%; Score 69; DB 10; Length 56;
Best Local Similarity 100.0%; Pred. No. 5.72e-05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 17 RYRPGTVAL 25
QY 1 RYRPGTVAL 9

RESULT 10
ID Q9XHN8 PRELIMINARY; PRT; 56 AA.
AC Q9XHN8;
DT 01-NOV-1999 (TRENBLrel. 12, Created)
DT 01-NOV-1999 (TRENBLrel. 12, Last sequence update)
DT 01-NOV-1999 (TRENBLrel. 12, Last annotation update)
DE HISTONE H3-D (FRAGMENT).
OS Glycine sp. G1545.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae;
OC Glycine.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CV. G1545;
RX MEDLINE: 99261647.
RA DOYLE J.J., DOYLE J.L., BROWN A.H.;
RT "Incongruence in the diploid B-genome species complex of Glycine
RT (Leguminosae) revisited: histone H3-D alleles versus chloroplast
RT haplotypes.";
RL Mol. Biol. Evol. 16:354-362(1999).
DR EMBL: AF093437; AAD40989.1; -.
FT NON_TER 1
FT NON_TER 56
SQ SEQUENCE 56 AA; 6506 MW; AA28B116 CRC32;

Query Match 100.0%; Score 69; DB 10; Length 56;
Best Local Similarity 100.0%; Pred. No. 5.72e-05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 17 RYRPGTVAL 25
QY 1 RYRPGTVAL 9

RESULT 11
ID Q9XHN2 PRELIMINARY; PRT; 56 AA.
AC Q9XHN2;
DT 01-NOV-1999 (TRENBLrel. 12, Created)
DT 01-NOV-1999 (TRENBLrel. 12, Last sequence update)
DT 01-NOV-1999 (TRENBLrel. 12, Last annotation update)
DE HISTONE H3-D (FRAGMENT).
OS Glycine sp. G2138.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae;
OC Glycine.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CV. G2138;
RX MEDLINE: 99261647.
RA DOYLE J.J., DOYLE J.L., BROWN A.H.;
RT "Incongruence in the diploid B-genome species complex of Glycine
RT (Leguminosae) revisited: histone H3-D alleles versus chloroplast
RT haplotypes.";
RL Mol. Biol. Evol. 16:354-362(1999).
DR EMBL: AF093437; AAD40989.1; -.
FT NON_TER 1
FT NON_TER 56
SQ SEQUENCE 56 AA; 6506 MW; AA28B116 CRC32;

Query Match 100.0%; Score 69; DB 10; Length 56;
Best Local Similarity 100.0%; Pred. No. 5.72e-05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 17 RYRPGTVAL 25
QY 1 RYRPGTVAL 9

RESULT 12
ID Q9XHN7 PRELIMINARY; PRT; 56 AA.
AC Q9XHN7;
DT 01-NOV-1999 (TRENBLrel. 12, Created)
DT 01-NOV-1999 (TRENBLrel. 12, Last sequence update)
DT 01-NOV-1999 (TRENBLrel. 12, Last annotation update)
DE HISTONE H3-D (FRAGMENT).
OS Glycine sp. G1077.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae;
OC Glycine.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CV. G1077;
RX MEDLINE: 99261647.
RA DOYLE J.J., DOYLE J.L., BROWN A.H.;
RT "Incongruence in the diploid B-genome species complex of Glycine
RT (Leguminosae) revisited: histone H3-D alleles versus chloroplast
RT haplotypes.";
RL Mol. Biol. Evol. 16:354-362(1999).
DR EMBL: AF093439; AAD40991.1; -.
FT NON_TER 1
FT NON_TER 56
SQ SEQUENCE 56 AA; 6506 MW; AA28B116 CRC32;

Query Match 100.0%; Score 69; DB 10; Length 56;
Best Local Similarity 100.0%; Pred. No. 5.72e-05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 17 RYRPGTVAL 25
QY 1 RYRPGTVAL 9

RESULT 13
ID Q42826 PRELIMINARY; PRT; 56 AA.
AC Q42826; Q42824; Q42827;
DT 01-NOV-1996 (TRENBLrel. 01, Created)
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
DT 01-NOV-1999 (TRENBLrel. 12, Last annotation update)
DE HISTONE H3 (FRAGMENT).
OS Glycine tabacina.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae;
OC Glycine.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 97131650.
RA DOYLE J.J., KANAZIN V., SHOEMAKER R.C.;
```

RT "Phylogenetic utility of histone H3 intron sequences in the perennial  
RL relatives of soybean (Glycine: Leguminosae).";  
RN Mol. Phylogenet. Evol. 6:438-447(1996).  
RP [2]

RA SEQUENCE FROM N.A.  
RP DOYLE J.J., KANAZIN V., SHOEMAKER R.C.;  
RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.  
RN [3]

RP SEQUENCE FROM N.A.  
RN STRAIN-CV. G2601, CV. G1138, CV. G1317, CV. G2260, CV. G2341;  
RC MEDLINE; 99261647.  
RX

RA DOYLE J.J., DOYLE J.L., BROWN A.H.;  
RN "Incongruence in the diploid B-genome species complex of Glycine  
RL (Leguminosae) revisited: histone H3-D alleles versus chloroplast  
RT haplotypes";  
RP Mol. Biol. Evol. 16:354-362(1999).  
RN

CC -1- FUNCTION: HISTONE H3, ALONG WITH HISTONE H4, PLAYS A CENTRAL ROLE  
CC IN NUCLEOSOME FORMATION.  
CC

CC -1- SUBUNIT: THE NUCLEOSOME IS AN OCTAMER CONTAINING TWO MOLECULES OF  
CC H2A, H2B, H3, AND H4; WHICH WRAP APPROXIMATIVELY 146 BP OF DNA.  
CC

CC EMBL; U47392; AAB50483.1; -;  
CC EMBL; U47374; AAB50445.1; -;  
CC EMBL; U47409; AAB50489.1; -;  
CC EMBL; AF0933452; AAD41004.1; -;  
CC EMBL; AF0933448; AAD41000.1; -;  
CC EMBL; AF0934449; AAD41001.1; -;  
CC EMBL; AF0934450; AAD41002.1; -;  
CC EMBL; AF093451; AAD41003.1; -;  
CC MENDEL; 15674; Glyta; His3;15674.  
KW Nuclear protein; Chromosomal protein; DNA-binding; Nucleosome core.  
FT NON\_TER 1  
FT SEQUENCE 56 AA; 56 MW; AA28B116 CRC32;  
SQ

Query Match 100.0%; Score 69; DB 10; Length 56;  
Best Local Similarity 100.0%; Pred. No. 5.72e-05;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 17 RYRPGTVAL 25  
| | | | | | | |  
QY 1 RYRPGTVAL 9

RESULT 14  
ID Q27899 PRELIMINARY; PRT; 83 AA.  
AC Q27899;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)  
DE HISTONE H3 (FRAGMENT).  
GN H3.3.

OS Leptothorax acervorum.  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata;  
OC Formicoidea; Formicidae; Leptothorax.  
RN [1]

RP SEQUENCE FROM N.A.  
RA BAUR A., STETZER N.E., BUSCHINGER A., ZIMMERMANN F.K.;  
RL Submitted (FEB-1994) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: HISTONE H3, ALONG WITH HISTONE H4, PLAYS A CENTRAL ROLE  
CC IN NUCLEOSOME FORMATION.  
CC

CC -1- SUBUNIT: THE NUCLEOSOME IS AN OCTAMER CONTAINING TWO MOLECULES OF  
CC H2A, H2B, H3, AND H4; WHICH WRAP APPROXIMATIVELY 146 BP OF DNA.  
CC

CC EMBL; X77741; CAA54791.1; -;  
CC EMBL; X77740; CAA54790.1; -;  
CC PFAM; PF00125; histone.1.  
KW Nuclear protein; Chromosomal protein; DNA-binding; Nucleosome core.  
FT NON\_TER 1  
FT SEQUENCE 83 AA; 9499 MW; F64CA6AC CRC32;  
SQ

Query Match 100.0%; Score 69; DB 5; Length 83;  
Best Local Similarity 100.0%; Pred. No. 5.72e-05;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 17 RYRPGTVAL 25  
| | | | | | | |  
QY 1 RYRPGTVAL 9

RESULT 14  
ID Q27899 PRELIMINARY; PRT; 83 AA.  
AC Q27899;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)  
DE HISTONE H3 (FRAGMENT).  
GN H3.3.

OS Leptothorax acervorum.  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata;  
OC Formicoidea; Formicidae; Leptothorax.  
RN [1]

RP SEQUENCE FROM N.A.  
RA BAUR A., STETZER N.E., BUSCHINGER A., ZIMMERMANN F.K.;  
RL Submitted (FEB-1994) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: HISTONE H3, ALONG WITH HISTONE H4, PLAYS A CENTRAL ROLE  
CC IN NUCLEOSOME FORMATION.  
CC

CC -1- SUBUNIT: THE NUCLEOSOME IS AN OCTAMER CONTAINING TWO MOLECULES OF  
CC H2A, H2B, H3, AND H4; WHICH WRAP APPROXIMATIVELY 146 BP OF DNA.  
CC

CC EMBL; X77741; CAA54791.1; -;  
CC EMBL; X77740; CAA54790.1; -;  
CC PFAM; PF00125; histone.1.  
KW Nuclear protein; Chromosomal protein; DNA-binding; Nucleosome core.  
FT NON\_TER 1  
FT SEQUENCE 83 AA; 9499 MW; F64CA6AC CRC32;  
SQ

Query Match 100.0%; Score 69; DB 5; Length 83;  
Best Local Similarity 100.0%; Pred. No. 5.72e-05;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 17 RYRPGTVAL 25  
| | | | | | | |  
QY 1 RYRPGTVAL 9

RESULT 14  
ID Q27899 PRELIMINARY; PRT; 83 AA.  
AC Q27899;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)  
DE HISTONE H3 (FRAGMENT).  
GN H3.3.

OS Leptothorax acervorum.  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata;  
OC Formicoidea; Formicidae; Leptothorax.  
RN [1]

RP SEQUENCE FROM N.A.  
RA BAUR A., STETZER N.E., BUSCHINGER A., ZIMMERMANN F.K.;  
RL Submitted (FEB-1994) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: HISTONE H3, ALONG WITH HISTONE H4, PLAYS A CENTRAL ROLE  
CC IN NUCLEOSOME FORMATION.  
CC

CC -1- SUBUNIT: THE NUCLEOSOME IS AN OCTAMER CONTAINING TWO MOLECULES OF  
CC H2A, H2B, H3, AND H4; WHICH WRAP APPROXIMATIVELY 146 BP OF DNA.  
CC

CC EMBL; X77741; CAA54791.1; -;  
CC EMBL; X77740; CAA54790.1; -;  
CC PFAM; PF00125; histone.1.  
KW Nuclear protein; Chromosomal protein; DNA-binding; Nucleosome core.  
FT NON\_TER 1  
FT SEQUENCE 83 AA; 9499 MW; F64CA6AC CRC32;  
SQ

Query Match 100.0%; Score 69; DB 5; Length 83;  
Best Local Similarity 100.0%; Pred. No. 5.72e-05;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 17 RYRPGTVAL 25  
| | | | | | | |  
QY 1 RYRPGTVAL 9

RESULT 14  
ID Q27899 PRELIMINARY; PRT; 83 AA.  
AC Q27899;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)  
DE HISTONE H3 (FRAGMENT).  
GN H3.3.

OS Leptothorax acervorum.  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata;  
OC Formicoidea; Formicidae; Leptothorax.  
RN [1]

RP SEQUENCE FROM N.A.  
RA BAUR A., STETZER N.E., BUSCHINGER A., ZIMMERMANN F.K.;  
RL Submitted (FEB-1994) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: HISTONE H3, ALONG WITH HISTONE H4, PLAYS A CENTRAL ROLE  
CC IN NUCLEOSOME FORMATION.  
CC

CC -1- SUBUNIT: THE NUCLEOSOME IS AN OCTAMER CONTAINING TWO MOLECULES OF  
CC H2A, H2B, H3, AND H4; WHICH WRAP APPROXIMATIVELY 146 BP OF DNA.  
CC

CC EMBL; X77741; CAA54791.1; -;  
CC EMBL; X77740; CAA54790.1; -;  
CC PFAM; PF00125; histone.1.  
KW Nuclear protein; Chromosomal protein; DNA-binding; Nucleosome core.  
FT NON\_TER 1  
FT SEQUENCE 83 AA; 9499 MW; F64CA6AC CRC32;  
SQ

Query Match 100.0%; Score 69; DB 5; Length 83;  
Best Local Similarity 100.0%; Pred. No. 5.72e-05;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 17 RYRPGTVAL 25  
| | | | | | | |  
QY 1 RYRPGTVAL 9

RESULT 14  
ID Q27899 PRELIMINARY; PRT; 83 AA.  
AC Q27899;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)  
DE HISTONE H3 (FRAGMENT).  
GN H3.3.

OS Leptothorax acervorum.  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata;  
OC Formicoidea; Formicidae; Leptothorax.  
RN [1]

RP SEQUENCE FROM N.A.  
RA BAUR A., STETZER N.E., BUSCHINGER A., ZIMMERMANN F.K.;  
RL Submitted (FEB-1994) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: HISTONE H3, ALONG WITH HISTONE H4, PLAYS A CENTRAL ROLE  
CC IN NUCLEOSOME FORMATION.  
CC

Db 8 RYRPGTVAL 16  
| | | | | | | |  
QY 1 RYRPGTVAL 9

RESULT 15  
ID P91947 PRELIMINARY; PRT; 90 AA.  
AC P91947;  
DT 01-MAY-1997 (TrEMBLrel. 03, Created)  
DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)  
DT 01-JAN-1998 (TrEMBLrel. 05, Last annotation update)  
DE HISTONE H3 (FRAGMENT).  
GN H3.

OS Drosophila virilis (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
RN [1]

RP SEQUENCE FROM N.A.  
RC STRAIN-BOCHUM;  
RA NAGEL S.;  
RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: HISTONE H3, ALONG WITH HISTONE H4, PLAYS A CENTRAL ROLE  
CC IN NUCLEOSOME FORMATION.  
CC

CC -1- SUBUNIT: THE NUCLEOSOME IS AN OCTAMER CONTAINING TWO MOLECULES OF  
CC H2A, H2B, H3, AND H4; WHICH WRAP APPROXIMATIVELY 146 BP OF DNA.  
CC

CC EMBL; U82928; AAB49448.1; -;  
CC FLYBASE; FBgn0013084; Dvir\His3.  
CC PROSITE; PS00959; HISTONE\_H3\_2; 1.  
CC PROSITE; PS00322; HISTONE\_H3\_1; 1.  
KW Nuclear protein; Chromosomal protein; DNA-binding; Nucleosome core.  
FT NON\_TER 90  
FT SEQUENCE 90 AA; 10222 MW; DE2A3A06 CRC32;  
SQ

Query Match 100.0%; Score 69; DB 5; Length 90;  
Best Local Similarity 100.0%; Pred. No. 5.72e-05;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 43 RYRPGTVAL 51  
| | | | | | | |  
QY 1 RYRPGTVAL 9

Search completed: Sat Apr 15 00:04:36 2000  
Job time : 103 secs.

**THIS PAGE BLANK (USPTO)**

\*\*\*\*\*  
[W][A][T][E][R][M][A][N]  
[M][O][D][E][L]  
[P][R][O][T][E][I][N]  
[S][E][Q][U][E][N][C][E]  
[A][N][A][L][Y][S][I][S]  
[S][O][F][T][W][A][R][E]  
[V][E][R][S][I][O][N]  
[1].0  
(TM)  
\*\*\*\*\*

Release 3.1A John F. Collins, Blocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Sat Apr 15 00:06:26 2000; MasPar time 5.58 Seconds  
Tabular output not generated. 38.189 Million cell updates/sec.

Title: >US-08-452-843-12  
Description: (1-9) from US08452843.pap  
Perfect Score: 69  
Sequence: 1 MPRGVVVTL 9

Scoring table: PAM 150  
Gap 15

Searched: 18963 seqs, 23686106 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: a-geneseq36  
1:geneseqp

Statistics: Mean 16.665; Variance 48.715; scale 0.342

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	69	100.0	9	1 R89373	B7 naturally processed	5.12e-01
2	51	73.9	49	1 W4862	Human secreted protein	6.32e-01
3	49	71.0	165	1 W38665	S. pneumoniae amino ac	1.05e-02
4	49	71.0	697	1 W46517	Saccharomyces cerevisi	1.05e-02
5	48	69.6	68	1 W9441	Staphylococcus aureus	1.35e-02
6	48	69.6	246	1 W8752	H. pylori GHPO 1051 pr	1.35e-02
7	48	69.6	337	1 W85047	A human 7-transmembran	1.35e-02
8	48	69.6	337	1 W5799	Human 7-transmembran	1.35e-02
9	48	69.6	561	1 W97413	Lucilia cuprina GABA r	1.35e-02
10	48	69.6	572	1 W8391	Aspergillus fumigatus	1.35e-02
11	48	69.6	1155	1 W5102	Mouse beta-integrin al	1.35e-02
12	48	69.6	1155	1 W60002	Mouse alpha d polypepti	1.35e-02
13	48	69.6	1155	1 W78167	Mouse alpha-d subunit.	1.35e-02
14	48	69.6	1155	1 W73346	Mouse alpha-d polypepti	1.35e-02
15	48	69.6	1155	1 W23060	Mouse beta 2 integrin	1.35e-02
16	48	69.6	1155	1 W72835	Mouse alpha-d #1.	1.35e-02
17	48	69.6	1161	1 W73347	Mouse alpha-d polypepti	1.35e-02
18	48	69.6	1161	1 W23061	Mouse beta 2 integrin	1.35e-02
19	48	69.6	1161	1 W78168	Mouse alpha-d subunit.	1.35e-02
20	48	69.6	1161	1 W60003	Mouse alpha d polypepti	1.35e-02
21	48	69.6	1161	1 W72836	Mouse alpha-d #2.	1.35e-02
22	48	69.6	1161	1 W5103	Mouse beta-integrin al	1.35e-02
23	47	68.1	97	1 W64300	Mycobacterium tubercul	1.74e-02

24	47	68.1	97	1 W32428	Mycobacterium tubercul	1.74e-02
25	47	68.1	97	1 W81663	M. tuberculosis immuno	1.74e-02
26	47	68.1	97	1 W32360	Mycobacterium tubercul	1.74e-02
27	47	68.1	151	1 W92413	Herpesvirus Saimiri OR	1.74e-02
28	47	68.1	151	1 W02387	HVS13 (viral homologue	1.74e-02
29	47	68.1	151	1 W13653	Herpesvirus Saimiri ORF	1.74e-02
30	47	68.1	151	1 W6571	Herpesvirus ORF13 prod	1.74e-02
31	47	68.1	229	1 R70701	Recombinant DNA-ase-B.	1.74e-02
32	47	68.1	271	1 R88702	Mitogenic factor assoc	1.74e-02
33	47	68.1	271	1 R88823	S. pyogenes DNaseB and	1.74e-02
34	47	68.1	293	1 R70702	DNA-ase-B	1.74e-02
35	47	68.1	464	1 Y04935	Mycobacterium species	1.74e-02
36	47	68.1	580	1 W32363	Mycobacterium tubercul	1.74e-02
37	47	68.1	580	1 W32431	Mycobacterium tubercul	1.74e-02
38	47	68.1	580	1 W81666	M. tuberculosis immuno	1.74e-02
39	47	68.1	580	1 W64303	Mycobacterium tubercul	1.74e-02
40	47	68.1	602	1 Y04996	Mycobacterium species	1.74e-02
41	47	68.1	2183	1 W48708	Measles virus Moraten	1.74e-02
42	47	68.1	2183	1 W48703	Measles virus Edmonsto	1.74e-02
43	47	68.1	2183	1 W48705	Measles virus 1983 iso	1.74e-02
44	47	68.1	2183	1 W48710	Measles virus AIR-C va	1.74e-02
45	47	68.1	2183	1 W48709	Measles virus Zagreb v	1.74e-02

ALIGNMENTS

RESULT 1  
ID R89373 standard; peptide; 9 AA.  
AC R89373; 1996 (first entry)  
DT 18-SEP-1996  
DE B7 naturally processed protein derived immunogenic peptide.  
KW Immunogenic peptide; supermotif; HLA molecule; CTL response;  
KW therapeutic; diagnostic; cancer; viral infection; hepatitis B;  
KW hepatitis C.  
OS Synthetic.  
PN W09603140-AL.  
PD 08-FEB-1996.  
PF 21-JUL-1995; U09234.  
PR 21-JUL-1994; US-278634.  
PR 23-NOV-1994; US-344824.  
PR 30-MAY-1995; US-452843.  
PA (CYTE-) CYTEL CORP.  
PI Sette A, Sidney J;  
DR WPI; 96-116784/12.  
PT Compsn. comprising immunogenic peptide with supermotif allowing more  
PT than one HLA mol. to bind - used to induce CTL response in patient  
PT and for in vivo and ex vivo therapeutic and diagnostic applications  
PS Claim 2; Page 26; 32pp; English.  
CC The sequences given in R89362-82 are immunogenic peptides which were  
CC use in the composition of the invention. The composition comprises  
CC an immunogenic peptide of 9-10 residues with a supermotif which  
CC allows binding of more than one HLA molecule. It pref. comprises  
CC two conserved residues, a first at the 2nd position from the N-  
CC terminal is Pro, and a 2nd at the C-terminal is Met. These peptides  
CC are used to induce a CTL response in a patient. They are also  
CC useful in compositions for in vivo and ex vivo therapeutic and  
CC diagnostic applications, e.g the treatment of cancer and viral  
CC infections, e.g. hepatitis B and C.  
SQ Sequence 9 AA;

Query Match 100.0%; Score 69; DB 1; Length 9;  
Best Local Similarity 100.0%; Pred. No. 5.12e-01;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 MPRGVVVTL 9  
| | | | | | | | | |  
Qy 1 MPRGVVVTL 9

RESULT 2  
ID W74862 standard; Protein; 49 AA.  
AC W74862;  
DT 19-JAN-1999 (first entry)

DE Human secreted protein encoded by gene 134 clone HPMGD24.  
KW Human; secreted protein; testis; tumour; foetal brain tissue;  
KW fusion protein; cancer; central nervous system; seizure;  
KW diagnosis; neurodegenerative disease.  
OS Homo sapiens.  
PN WO9839448-A2.  
PD 11-SEP-1998.  
PF 06-MAR-1998; U04493.  
PR 02-OCT-1997; US-061060.  
PR 07-MAR-1997; US-038621.  
PR 07-MAR-1997; US-040161.  
PR 07-MAR-1997; US-040162.  
PR 07-MAR-1997; US-040163.  
PR 07-MAR-1997; US-040333.  
PR 07-MAR-1997; US-040334.  
PR 07-MAR-1997; US-040336.  
PR 07-MAR-1997; US-040626.  
PR 11-APR-1997; US-043311.  
PR 11-APR-1997; US-043312.  
PR 11-APR-1997; US-043313.  
PR 11-APR-1997; US-043314.  
PR 11-APR-1997; US-043568.  
PR 11-APR-1997; US-043569.  
PR 11-APR-1997; US-043576.  
PR 11-APR-1997; US-043578.  
PR 11-APR-1997; US-043580.  
PR 11-APR-1997; US-043669.  
PR 11-APR-1997; US-043670.  
PR 11-APR-1997; US-043671.  
PR 11-APR-1997; US-043672.  
PR 11-APR-1997; US-043674.  
PR 23-MAY-1997; US-047492.  
PR 23-MAY-1997; US-047500.  
PR 23-MAY-1997; US-047501.  
PR 23-MAY-1997; US-047502.  
PR 23-MAY-1997; US-047503.  
PR 23-MAY-1997; US-047581.  
PR 23-MAY-1997; US-047582.  
PR 23-MAY-1997; US-047583.  
PR 23-MAY-1997; US-047584.  
PR 23-MAY-1997; US-047585.  
PR 23-MAY-1997; US-047586.  
PR 23-MAY-1997; US-047587.  
PR 23-MAY-1997; US-047588.  
PR 23-MAY-1997; US-047589.  
PR 23-MAY-1997; US-047590.  
PR 23-MAY-1997; US-047592.  
PR 23-MAY-1997; US-047593.  
PR 23-MAY-1997; US-047594.  
PR 23-MAY-1997; US-047595.  
PR 23-MAY-1997; US-047596.  
PR 23-MAY-1997; US-047597.  
PR 23-MAY-1997; US-047598.  
PR 23-MAY-1997; US-047599.  
PR 23-MAY-1997; US-047600.  
PR 23-MAY-1997; US-047601.  
PR 23-MAY-1997; US-047612.  
PR 23-MAY-1997; US-047613.  
PR 23-MAY-1997; US-047614.  
PR 23-MAY-1997; US-047615.  
PR 23-MAY-1997; US-047617.  
PR 23-MAY-1997; US-047618.  
PR 23-MAY-1997; US-047632.  
PR 23-MAY-1997; US-047633.  
PR 06-JUN-1997; US-048954.  
PR 06-JUN-1997; US-048974.  
PR 13-JUN-1997; US-049610.  
PR 08-JUL-1997; US-051926.  
PR 16-JUL-1997; US-052874.  
PR 18-AUG-1997; US-055724.  
PR 22-AUG-1997; US-056630.  
PR 22-AUG-1997; US-056631.  
PR 22-AUG-1997; US-056632.  
PR 22-AUG-1997; US-056636.  
PR 22-AUG-1997; US-056637.  
PR 22-AUG-1997; US-056662.  
PR 22-AUG-1997; US-056664.  
PR 22-AUG-1997; US-056845.  
PR 22-AUG-1997; US-056862.  
PR 22-AUG-1997; US-056864.  
PR 22-AUG-1997; US-056872.  
PR 22-AUG-1997; US-056874.  
PR 22-AUG-1997; US-056875.  
PR 22-AUG-1997; US-056876.  
PR 22-AUG-1997; US-056877.  
PR 22-AUG-1997; US-056878.  
PR 22-AUG-1997; US-056879.  
PR 22-AUG-1997; US-056880.  
PR 22-AUG-1997; US-056881.  
PR 22-AUG-1997; US-056882.  
PR 22-AUG-1997; US-056884.  
PR 22-AUG-1997; US-056886.  
PR 22-AUG-1997; US-056887.  
PR 22-AUG-1997; US-056888.  
PR 22-AUG-1997; US-056889.  
PR 22-AUG-1997; US-056892.  
PR 22-AUG-1997; US-056893.  
PR 22-AUG-1997; US-056894.  
PR 22-AUG-1997; US-056903.  
PR 22-AUG-1997; US-056908.  
PR 22-AUG-1997; US-056909.  
PR 22-AUG-1997; US-056910.  
PR 22-AUG-1997; US-056911.  
PR 05-SEP-1997; US-057850.  
PR 05-SEP-1997; US-057669.  
PR 05-SEP-1997; US-057761.  
PR 12-SEP-1997; US-058785.  
PR (HUMA-) HUMAN GENOME SCI INC.  
PI Bednarik DP, Brewer LA, Carter KC, Duan R, Ebner R, Endress GA,  
PI Feng P, Ferrie AM, Fischer CL, Florence KA, Greene JM, Hu JS,  
PI Kyaw H, Lafleur DW, Li Y, Moore PA, Ni J, Olsen HS, Rosen CA,  
PI Ruben SM, Shi Y, Soppet DR, Young PE, Yu GL, Zeng Z;  
PI WPI: 98-506364/43.  
PI N-PSDB: V59644.  
PT New isolated human genes and the secreted polypeptide(s) they encode  
PT - useful for diagnosis and treatment of e.g. cancers, neurological  
PT disorders, immune diseases, inflammation or blood disorders  
PS Claim 1: Page 619; 721pp; English.  
CC This sequence represents a secreted human protein encoded by the nucleic  
CC acid molecule designated Gene 134 from the human cDNA clone HPMGD24  
CC (deposited as clone ATCC 97902 and ATCC 209048).  
CC The gene can be used to generate fusion proteins by linking to the gene  
CC to a human immunoglobulin Fc portion (e.g. V59502) for increasing the  
CC stability of the fused protein as compared to the human protein only.  
CC The invention relates to 186 novel genes and their fragments (nucleic  
CC acid sequences: V59511-V59812; amino acid sequences W74731-W75026) which  
CC are useful for preventing, treating or ameliorating medical conditions  
CC e.g. by protein or gene therapy. Also, pathological conditions can be  
CC diagnosed by determining the amount of the new polypeptides in a sample  
CC or by determining the presence of mutations in the new polynucleotides.  
CC Specific uses are described for each of the 186 polynucleotides, based on  
CC which tissues they are most highly expressed in (see V59511 for described  
CC uses).  
SQ Sequence 49 AA;  
Query Match 73.9%; Score 51; DB 1; Length 49;  
Best Local Similarity 75.0%; Pred. NO. 6.32e+01;  
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
Db 4 LPRGVVVS 11  
QY 1 MPRGVVVT 8  
RESULT 3  
ID W38665 standard; Protein: 165 AA.

AC W38665;  
 DT 09-NOV-1998 (first entry)  
 DE S. pneumoniae amino acid permease ROCE.  
 KW Streptococcus pneumoniae protein; genetic immunisation; antagonist;  
 KW immunological response; inoculation; antibody production; inhibitor;  
 KW T cell immune response; antimicrobial compound; bacterial adhesion;  
 KW extracellular matrix protein; protein-mediated cell invasion; wound;  
 KW pathogenesis.  
 OS Streptococcus pneumoniae.  
 PN WO9743303-A1.  
 PD 20-NOV-1997.  
 PF 14-MAY-1997; U07950.  
 PR 14-MAY-1996; US-017670.  
 PA (SMIK ) SMITHKLINE BEECHAM CORP.  
 PA (SMIK ) SMITHKLINE BEECHAM PLC.  
 PI Black MT, Hodgson JE, Knowles DJC, Nicholas RO,  
 PI Stodola RK;  
 DR WPI; 98-008793/01.  
 DR N-PSDB; T98709.  
 DT Novel Streptococcus pneumoniae proteins and related DNA - useful for  
 PT diagnosing anti-microbial agents for treatment of bacterial  
 PT infections  
 PS Claim 12: Pages 414-415; 483pp; English.  
 CC This sequence represents a Streptococcus pneumoniae protein that, based  
 CC on homology with a bacillus subtilis protein, is an amino acid permease  
 CC ROCE, and is encoded by a DNA sequence of the invention.  
 CC The DNA sequences were isolated from Streptococcus pneumoniae strain  
 CC 0100993 (NCIMB 40794). The Streptococcus pneumoniae proteins of the  
 CC invention can be used to identify compounds which interact with and  
 CC inhibit or activate the activity of the proteins. Antagonists can be  
 CC used to treat diseases caused by S. pneumoniae proteins, through genetic  
 CC immunisation. They can also be used to induce an immunological response  
 CC in a mammal by inoculation with the S. pneumoniae proteins or delivery  
 CC of the encoding nucleic acids in a vector adequate to produce antibody  
 CC and/or T cell immune responses to protect the animal from disease. The  
 CC proteins can also be used to identify antimicrobial compounds which are  
 CC capable of inhibiting their bioactivity. In particular the proteins of  
 CC the invention can be used to prevent adhesion of bacteria to mammalian  
 CC extracellular matrix proteins on in-dwelling devices or in wounds, to  
 CC block protein-mediated mammalian cell invasion, and to block the normal  
 CC progression of pathogenesis in infections initiated other than by the  
 CC implantation of in-dwelling devices or other surgical techniques.  
 SQ Sequence 165 AA;

Query Match 71.0%; Score 49; DB 1; Length 165;  
 Best Local Similarity 55.6%; Pred. No. 1.05e+02;  
 Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 33 IPRGIVLSL 41  
 QY :|||:|:|  
 1 MPRGVVVT 9

RESULT 4  
 ID W46517 standard; Protein; 697 AA.  
 AC W46517;  
 DT 23-JUN-1998 (first entry)  
 DE Saccharomyces cerevisiae V1a viral capsid-polymerase fusion protein.  
 KW Saccharomyces cerevisiae V1l viral capsid-polymerase fusion protein;  
 KW Scv1a totivirus; Scv1l.  
 OS Saccharomyces cerevisiae.  
 FH Key Location/Qualifiers  
 FT 202..442  
 FT Domain /label= "Multimerisation domain"  
 PN WO9800525-A1.  
 PD 08-JAN-1998.  
 PF 26-JUN-1997; U11216.  
 PR 02-JUL-1996; US-674351.  
 PA (UYNY ) UNIV NEW YORK STATE RES FOUND.  
 PI Bruenn JA, Yao W;  
 DR WPI; 98-086952/08.  
 DR N-PSDB; V05285.  
 DT Viral capsid polypeptide capable of inhibiting viral packaging -

PT comprises part of viral capsid proteins, useful in, e.g. recombinant  
 PT protein production of totiviruses  
 PS Claim 9; Pages 43-44; 52pp; English.  
 CC The present sequence represents a capsid-polymerase fusion protein  
 CC of Saccharomyces cerevisiae virus 1a (Scv1a). The virus is a member of  
 CC the totiviruses in which all the viral functions are encoded by a single  
 CC double strand of DNA. Scv1a contains two overlapping open reading  
 CC frames encoding a capsid protein and a polymerase protein (see V05285).  
 CC The latter is produced by translational frameshifting. The capsid  
 CC protein has been shown to inhibit viral packaging of its cognate virus  
 CC in yeast. The invention relates to the use of the sequence encoding the  
 CC capsid protein (amino acids 1-443 of this sequence) for conferring  
 CC resistance in yeast to the Scv1l and Scv1a totiviruses which can infect  
 CC yeast cultures especially those that are used to produce recombinant  
 CC proteins. Expression of the capsid proteins in other hosts e.g. plants  
 CC or animals, can also be used to inhibit virus packaging thereby  
 CC preventing viral spread and further infection.  
 SQ Sequence 697 AA;

Query Match 71.0%; Score 49; DB 1; Length 697;  
 Best Local Similarity 55.6%; Pred. No. 1.05e+02;  
 Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 393 MNRGIIIVDL 401  
 QY | :|:|:|  
 1 MPRGVVVT 9

RESULT 5  
 ID W79441 standard; Protein; 68 AA.  
 AC W79441;  
 DT 04-DEC-1998 (first entry)  
 DE Staphylococcus aureus protein.  
 KW Mycobacterium tuberculosis; treatment; prevention; bacterial infection;  
 KW Helicobacter pylori; vaccine.  
 OS Staphylococcus aureus.  
 PN WO9823738-A2.  
 PD 04-JUN-1998.  
 PF 24-NOV-1997; U22092.  
 PR 25-NOV-1996; US-031469.  
 PA (SMIK ) SMITHKLINE BEECHAM CORP.  
 PI Warren RL;  
 DR WPI; 98-322718/28.  
 DR N-PSDB; V59914.  
 DT New nucleic acid from Staphylococcus aureus NCIMB 40771 - useful  
 PT for, e.g. diagnosis, prevention and treatment of bacterial  
 PT infection(s)  
 PS Claim 5; Page 94; 114pp; English.  
 CC W79441-43 represent Staphylococcus aureus WCHU (NCIMB 40771)  
 CC proteins that have homology to a Mycobacterium tuberculosis  
 CC protein of unknown function. The S. aureus proteins are used to  
 CC generate antibodies and to screen for antimicrobials. The products  
 CC are used to treat or prevent bacterial infections, particularly  
 CC where caused by S. aureus but also against Helicobacter pylori.  
 CC Particular applications are to treat subjects before surgery or  
 CC insertion of an in-dwelling device (alternatively the device itself  
 CC is impregnated before placement). The nucleic acid sequence is used  
 CC as sources of antisense sequences (for therapeutic use) or  
 CC regulatory elements for controlling expression of bacterial genes,  
 CC and for antibacterial screening. The protein can be also used as a  
 CC vaccine.  
 SQ Sequence 68 AA;

Query Match 69.6%; Score 48; DB 1; Length 68;  
 Best Local Similarity 55.6%; Pred. No. 1.35e+02;  
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 36 MPRGILGSL 44  
 QY | :|:|:|  
 1 MPRGVVVT 9

RESULT 6

ID W98752 standard; Protein; 246 AA.  
AC W98752;  
DE H. pylori GHPO 1051 protein.  
KW GHPO protein; Helicobacter infection; gastroduodenal disease; gastritis;  
KW peptic ulcer disease.  
OS Helicobacter pylori.  
PN W09843478-A1.  
PD 08-OCT-1998.  
PF 01-APR-1998; U06371.  
PR 29-JUL-1997; US-902615.  
PR 01-APR-1997; US-833457.  
PR 24-JUN-1997; US-881227.  
PA (HUMA-) HUMAN GENOME SCI INC.  
PA (INMR) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.  
PI Al-Garawi A, Kleanthous H, Miller C, Oomen RP, Tomb J;  
DR WPI; 98-542293/46.  
DR N-PSDB; X14471.  
DE New isolated Helicobacter polynucleotides - used to develop products  
PT for the diagnosis, prevention and treatment of Helicobacter  
PT infections and gastrointestinal diseases  
PS Claim 8; Page 1608-1609; 2054pp; English.  
CC This sequence represents a Helicobacter pylori GHPO protein of the  
CC invention. The polypeptides can be used for preventing or treating  
CC Helicobacter infections, and gastroduodenal diseases associated with  
CC these infections, including acute, chronic, and atrophic gastritis, and  
CC peptic ulcer diseases, e.g. gastric and duodenal ulcers. They can also be  
CC used for the production of antibodies. The products can also be used for  
CC detection and diagnosis.  
CC Sequence 246 AA;  
SQ

Query Match 69.6%; Score 48; DB 1; Length 246;  
Best Local Similarity 66.7%; Pred. No. 1.35e-02;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
Db 155 KPGVLVIL 163  
QY 1 MPRGVVVT 9

RESULT 7  
ID W85047 standard; Protein; 337 AA.  
AC W85047;  
DE 05-FEB-1999 (first entry)  
DE A human 7-transmembrane receptor designated HMTMF81.  
KW Human; 7-transmembrane receptor protein; HMTMF81; infection;  
KW HIV; pain; cancer; anorexia; bulimia; asthma; Parkinson's disease;  
KW acute heart failure; hypotension; hypertension; urinary retention;  
KW osteoporosis; angina pectoris; myocardial infarction; ulcer; asthma;  
KW benign prostatic hypertrophy; neurological disorder.  
OS Homo sapiens.  
PN EP-878542-A2.  
PD 18-NOV-1998.  
PF 27-OCT-1997; 308560.  
PR 22-APR-1997; US-844795.  
PA (SMIK) SMITHKLINE BEECHAM CORP.  
PI Ellis CE, Halsey WS, Sathe GM;  
DR WPI; 98-585747/50.  
DR N-PSDB; V71117.  
PT DNA encoding 7-transmembrane receptor polypeptide HMTMF81 - useful  
PT for treatment of, e.g. HIV infections, pain, cancers, myocardial  
PT infarction and acute heart failure  
PS Claim 1; Pages 18-19; 20pp; English.  
CC The present sequence represents a human 7-transmembrane receptor protein  
CC designated HMTMF81. HMTMF81 polypeptides and polynucleotides can be  
CC used in the treatment of infections such as bacterial, fungal,  
CC protozoan and viral infections, particularly infections caused by  
CC HIV-1 or HIV-2. They can also be used to treat pain, cancers, anorexia,  
CC bulimia, asthma, Parkinson's disease, acute heart failure, hypotension,  
CC hypertension, urinary retention, osteoporosis, angina pectoris,  
CC myocardial infarction, ulcers, asthma, allergies, benign prostatic  
CC hypertrophy and psychotic and neurological disorders, including anxiety,  
CC schizophrenia, manic depression, delirium, dementia, severe mental

Query Match 69.6%; Score 48; DB 1; Length 337;  
Best Local Similarity 55.6%; Pred. No. 1.35e-02;  
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;  
Db 273 MOKSVVITL 281  
QY 1 MPRGVVVT 9

RESULT 8  
ID W75799 standard; Protein; 337 AA.  
AC W75799;  
DE Human 7-transmembrane receptor polypeptide, HMTMF81.  
KW HMTMF81; 7-transmembrane receptor; treatment; infection; bacteria; pain;  
KW fungal; protozoan; viral; human immune deficiency virus; HIV-1; HIV-2;  
KW cancer; anorexia; bulimia; asthma; Parkinson's disease; heart failure;  
KW hypotension; hypertension; urinary retention; osteoporosis; allergy;  
KW angina pectoris; myocardial infarction; asthma; ulcer; anxiety;  
KW prostatic hypertrophy; psychotic disorder; neurological disorder; human;  
KW schizophrenia; manic depression; delirium; dementia; mental retardation;  
KW dyskinesias; Huntington's disease; Gilles de la Tourette's syndrome.  
OS Homo sapiens.  
PN EP-874047-A2.  
PD 28-OCT-1998.  
PF 20-APR-1998; 303008.  
PR 19-MAR-1998; US-844795.  
PR 22-APR-1997; US-844795.  
PA (SMIK) SMITHKLINE BEECHAM CORP.  
PA (SMIK) SMITHKLINE BEECHAM PLC.  
PI Ames R, Chambers J, Ellis C, Foley J, Halsey W,  
PI Sarau H, Sathe G;  
DR WPI; 98-544641/47.  
DR N-PSDB; V62388.  
PT DNA encoding 7-trans-membrane receptor polypeptide HMTMF81 - useful  
PT in treatment of e.g. infections such as bacterial, fungal, protozoan  
PT and viral infections, particularly HIV, cancers and bulimia etc.  
PS Claim 11; Pages 7-8; 22pp; English.  
CC This represents a human 7-transmembrane receptor polypeptide, HMTMF81.  
CC The HMTMF81 polypeptides and polynucleotides can be used in the treatment  
CC of infections such as bacterial, fungal, protozoan and viral infections.  
CC They can be used particularly for treatment of infections caused by  
CC (human immune deficiency virus) HIV-1 or HIV-2, pain, cancers, anorexia,  
CC bulimia, asthma, Parkinson's disease, acute heart failure, hypotension,  
CC hypertension, urinary retention, osteoporosis, angina pectoris,  
CC myocardial infarction, ulcers, asthma, allergies, benign prostatic  
CC hypertrophy, and psychotic and neurological disorders, including  
CC anxiety, schizophrenia, manic depression, delirium, dementia, severe  
CC mental retardation and dyskinesias such as Huntington's disease or  
CC Gilles de la Tourette's syndrome.  
CC Sequence 337 AA;  
SQ

Query Match 69.6%; Score 48; DB 1; Length 337;  
Best Local Similarity 55.6%; Pred. No. 1.35e-02;  
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;  
Db 273 MOKSVVITL 281  
QY 1 MPRGVVVT 9

RESULT 9  
ID W97413 standard; Protein; 561 AA.  
AC W97413;  
DE 19-MAY-1999 (first entry)  
DE Lucilia cuprina GABA receptor subunit.  
KW Gamma-aminobutyric acid receptor; GABA receptor; sheep blow fly;  
KW pesticide.  
OS Lucilia cuprina.



PN RD-403074-A.  
PD 10-NOV-1997.  
PF 20-OCT-1997; 403074.  
PR 20-OCT-1997; RD-403074.  
PA (ANON ) ANONYMOUS.  
DR WPI: 98-007412/01.  
DR N-PSDB; X16065.  
PT GABA receptor sub-unit from L. cuprina - used in the production of  
PT effective pesticides.  
PS Disclosure; Fig 1; 2pp; English.  
CC The present sequence represents a gamma-aminobutyric acid (GABA)  
CC receptor subunit from the sheep blow fly, *Lucilia cuprina*. The GABA  
CC receptor subunit protein may be used to develop novel pesticides.  
SQ Sequence 561 AA;  
Query Match 69.6%; Score 48; DB 1; Length 561;  
Best Local Similarity 55.6%; Pred. No. 1.35e+02;  
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;  
Db 14 MPRSRIITL 22  
QY 1 MPRGVVVTL 9  
RESULT 10  
ID W69391 standard; Protein; 572 AA.  
AC W69391;  
DT 24-DEC-1998 (first entry)  
DE Aspergillus fumigatus protein 2.  
KW ss; auxotrophic cell line; histidine; adenyllic acid; leucine; growth;  
KW reproduction; antimicrobial.  
OS Aspergillus fumigatus.  
PN WO9841621-A1.  
PD 24-SEP-1998.  
PF 18-MAR-1998; U05350.  
PR 18-MAR-1997; US-041300.  
PA (MILL-) MILLENNIUM PHARM INC.  
PI Gavria V;  
DR WPI: 98-521216/44.  
DR N-PSDB; V58698.  
PT Aspergillus fumigatus polynucleotide(s) important for growth and  
PT reproduction - and auxotroph(s) created by omitting  
PT polynucleotide(s), useful e.g. to test functionality of unknown  
PT function cDNA from A. fumigatus cDNA library  
PS Claim 1; Fig 2; 53pp; English.  
CC The omission of one or more polynucleotides from Aspergillus fumigatus  
CC results in auxotrophic cell lines, which require media supplemented with  
CC histidine, adenyllic acid or leucine respectively for growth and  
CC reproduction. The auxotrophic cell lines (especially which require  
CC histidine, adenyllic acid or leucine to grow and reproduce) and  
CC polynucleotides can be used to test the functionality of unknown function  
CC cDNA from an A. fumigatus cDNA library. The polynucleotides and  
CC polypeptides are useful to identify agonists which may enhance growth  
CC and/or reproduction of A. fumigatus e.g. in the fermentation industry;  
CC they may also be administered (e.g. by inclusion of multiple gene copies)  
CC to enhance such growth and/or reproduction. The polypeptides can be used  
CC to produce antibodies, useful to detect polypeptides, screen for similar  
CC polypeptides from other organisms and as antimicrobials.  
SQ Sequence 572 AA;  
Query Match 69.6%; Score 48; DB 1; Length 572;  
Best Local Similarity 75.0%; Pred. No. 1.35e+02;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
Db 505 MPRGVVPA 512  
QY 1 MPRGVVVTL 8  
RESULT 11  
ID W65102 standard; Protein; 1155 AA.  
AC W65102;  
DT 28-SEP-1998 (first entry)

DE Mouse beta-integrin alpha-d putative human homologue protein.  
KW Beta-integrin alpha-d subunit; modulator; treatment; psoriasis;  
KW type-I diabetes; atherosclerosis; multiple sclerosis; asthma; murine;  
KW lung inflammation; acute respiratory distress syndrome;  
KW rheumatoid arthritis.  
OS Mus sp.  
PN US5728533-A.  
PD 17-MAR-1998.  
PF 07-JUN-1995; 485618.  
PR 23-DEC-1993; US-485618.  
PR 23-DEC-1993; US-173497.  
PR 05-AUG-1994; US-286889.  
PR 21-DEC-1994; US-362652.  
PA (ICOS-) ICOS CORP.  
PI Gallatin WM, Van Der Vieren M;  
DR WPI: 98-206565/18.  
DR N-PSDB; V35267.  
PT Screening assay for modulators of integrin binding - using  
PT immobilised or labelled alpha-d polypeptide, useful for, e.g.  
PT treating type-I diabetes  
PS Example 19; Column 115-122; 106pp; English.  
CC This sequence represents a mouse beta-integrin alpha-d subunit which is a  
CC putative homologue of the human sequence. This subunit is used in a  
CC method for identifying compounds that modulate the interaction of alpha-d  
CC with a binding partner of alpha-d which involves contacting an alpha-d  
CC polypeptide with an alpha-d binding partner, one of which is immobilised  
CC and the other of which is labelled, in the presence of a test compound,  
CC and determining if the compound affects binding between the alpha-d  
CC polypeptide and alpha-d binding partner, where the alpha-d polypeptide is  
CC alpha-d or its fragment comprising the cytoplasmic, transmembrane or  
CC extracellular domain of alpha-d. Compounds that modulate alpha-d binding  
CC could be used to treat diseases such as type-I diabetes, atherosclerosis,  
CC multiple sclerosis, asthma, psoriasis, lung inflammation, acute  
CC respiratory distress syndrome and rheumatoid arthritis.  
SQ Sequence 1155 AA;

Query Match 69.6%; Score 48; DB 1; Length 1155;  
Best Local Similarity 66.7%; Pred. No. 1.35e+02;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
Db 1 MVRGVVVILL 9  
QY 1 MPRGVVVTL 9  
RESULT 12  
ID W60002 standard; Protein; 1155 AA.  
AC W60002;  
DT 24-AUG-1998 (first entry)  
DE Mouse alpha d polypeptide.  
KW Beta 2 integrin alpha subunit; alpha d; human; treatment; diabetes;  
KW reporter-transactivator construct; arteriosclerosis; atherosclerosis;  
KW inflammatory bowel disease; arthritis; multiple sclerosis; mouse.  
OS Mus sp.  
PN US5766850-A.  
PD 16-JUN-1998.  
PF 21-DEC-1994; 362652.  
PR 21-DEC-1994; US-362652.  
PR 23-DEC-1993; US-173497.  
PR 05-AUG-1994; US-286889.  
PA (ICOS-) ICOS CORP.  
PI Gallatin WM, Van Der Vieren M;  
DR WPI: 98-361678/31.  
DR N-PSDB; V31569.  
PT Isolation of DNA encoding protein that binds to integrin subunit -  
PT using recombinant cells containing reporter-transactivator construct  
PS Example 19; Columns 99-106; 86pp; English.  
CC This represents a mouse cDNA clone encoding an alpha d subunit having  
CC homology to the human beta2 integrin alpha subunit (alpha d) subunit. The  
CC invention provides methods for isolating a polynucleotide encoding a  
CC protein that binds to alpha d. The method comprises transforming or  
CC transfecting host cells with a DNA construct comprising a reporter gene  
CC under the control of a promoter regulated by a transcription factor

CC having a DNA-binding domain and an activating domain. A first hybrid DNA  
 CC sequence encoding a fusion of at least part of alpha d and either the  
 CC DNA-binding domain or the activating domain of the transcription factor  
 CC and a library of hybrid DNA sequences encoding fusions of at least part  
 CC of putative alpha d-binding proteins and the DNA-binding domain or the  
 CC activating domain of the transcription factor which is not incorporated  
 CC in the first fusion are expressed in the host cells. The binding of an  
 CC alpha d-binding protein to alpha d in a particular host cell is detected  
 CC by determining production of the reporter gene product in the cell. The  
 CC hybrid DNA sequence encoding the alpha d-binding protein can be isolated  
 CC from the cell. Alpha d may be useful for treating graft arteriosclerosis,  
 CC atherosclerosis, diabetes, inflammatory bowel disease, arthritis and  
 CC multiple sclerosis.  
 SQ Sequence 1155 AA;

Query Match 69.6%; Score 48; DB 1; Length 1155;  
 Best Local Similarity 66.7%; Pred. No. 1.35e+02;  
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 1 MVRGVVILL 9  
 | | | | |  
 QY 1 MPRGVVVTL 9

## RESULT 13

ID R78167 standard; Protein; 1155 AA.  
 AC R78167;  
 DT 28-DEC-1995 (first entry)  
 DE Mouse alpha-d subunit.  
 KW Beta-2 integrin alpha-d subunit; antiinflammatory; arteriosclerosis;  
 KW inflammatory bowel disease; asthma; knock-out mouse.  
 OS Mus sp.  
 PN WO9517412-A1.  
 PD 29-JUN-1995.  
 PF 21-DEC-1994; U14832.  
 PR 23-DEC-1993; US-173497.  
 PR 05-AUG-1994; US-286889.  
 PA (ICOS-) ICOS CORP.  
 PI Gallatin WM, Van Der Vieren M;  
 DR WPI: 95-240603/31.  
 DR N-PSDB; Q91713.

PT Alpha sub-unit polypeptide of human beta 2 integrin - used to  
 PT identify potential antiinflammatory agents, for the treatment of  
 PT graft arteriosclerosis, inflammatory bowel disease, asthma, etc.  
 PS Disclosure; Page 118-123; 172pp; English.  
 CC A probe based on human integrin alpha-d clone 19A2 (given in  
 CC Q91712) was used to isolate mouse alpha-d cDNA clones from a thymic  
 CC oligo dt-primed library in lambda ZAP II. RACE PCR was used  
 CC to obtain a composite sequence (Q91713) encoding a putative  
 CC mouse alpha-d clone.  
 SQ Sequence 1155 AA;

Query Match 69.6%; Score 48; DB 1; Length 1155;  
 Best Local Similarity 66.7%; Pred. No. 1.35e+02;  
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 1 MVRGVVILL 9  
 | | | | |  
 QY 1 MPRGVVVTL 9

## RESULT 14

ID W73346 standard; Protein; 1155 AA.  
 AC W73346;  
 DT 11-FEB-1999 (first entry)  
 DE Mouse alpha d protein sequence.  
 KW Beta integrin subunit alpha d; VCAM-1; vascular cell adhesion molecule-1;  
 KW modulator identification; type I diabetes; atherosclerosis; psoriasis;  
 KW multiple sclerosis; asthma; lung inflammation; rheumatoid arthritis;  
 KW adult respiratory distress syndrome; therapy.  
 OS Mus sp.  
 PN U55837478-A.  
 PD 17-NOV-1998.

PF 03-OCT-1997; 943363.  
 PR 03-OCT-1997; US-943363.  
 PR 23-DEC-1993; US-173497.  
 PR 05-AUG-1994; US-286889.  
 PR 21-DEC-1994; US-362652.  
 PR 22-FEB-1996; US-605672.  
 PA (ICOS-) ICOS CORP.  
 PI Gallatin WM, Van Der Vieren M;  
 DR WPI: 99-023443/02.  
 DR N-PSDB; V08486.  
 PT Identifying modulators of interaction between alpha-d subunit of  
 PT human beta-2 integrin and vascular cell adhesion molecule-1 - useful  
 PT for treating e.g. diabetes, atherosclerosis, asthma and rheumatoid  
 PT arthritis  
 PS Example 28; Column 145-153; 128pp; English.  
 CC This sequence is the mouse beta integrin subunit alpha d (I) and can  
 CC be used in the method of the invention. The method is for the  
 CC identification of modulators (A) of binding between the (I) and VCAM-1  
 CC (vascular cell adhesion molecule-1), and comprises: (a) combining (I) and  
 CC VCAM-1 in presence of a test compound; and (b) detecting any decrease or  
 CC increase in binding between (I) and VCAM-1. (A) are potentially useful  
 CC for treating diseases associated with binding of (I) to its ligands,  
 CC e.g. type I diabetes, atherosclerosis, multiple sclerosis, asthma,  
 CC psoriasis, lung inflammation, adult respiratory distress syndrome and  
 CC rheumatoid arthritis. The method may also be modified to isolate a  
 CC nucleic acid that encodes (A), e.g. in a two-hybrid assay.  
 SQ Sequence 1155 AA;

Query Match 69.6%; Score 48; DB 1; Length 1155;  
 Best Local Similarity 66.7%; Pred. No. 1.35e+02;  
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 1 MVRGVVILL 9  
 | | | | |  
 QY 1 MPRGVVVTL 9

## RESULT 15

ID W23060 standard; Protein; 1155 AA.  
 AC W23060;  
 DT 24-FEB-1998 (first entry)  
 DE Mouse beta 2 integrin alpha d subunit.  
 KW Beta 2 integrin alpha d subunit; mouse; cell migration;  
 KW cell adhesion; phagocytosis; diabetes; atherosclerosis;  
 KW multiple sclerosis; asthma; psoriasis; lung inflammation;  
 KW acute respiratory distress syndrome; rheumatoid arthritis;  
 KW monoclonal antibody.  
 OS Mus musculus.  
 PN WO9731099-A1.  
 PD 28-AUG-1997.  
 PR 22-FEB-1997; U02713.  
 PR 24-FEB-1996; US-605672.  
 PA (ICOS-) ICOS CORP.  
 PI Gallatin WM, Van Der Vieren M;  
 DR WPI: 97-435154/40.  
 DR N-PSDB; T79251.  
 PT Hybridoma 199M and antibody secreted by it - specific for new rat  
 PT beta2 integrin subunit, useful to detect subunit in cells and  
 PT modulate its activity  
 PS Example 19; Page 150-155; 222pp; English.  
 CC This polypeptide comprises a murine homologue of a novel human  
 CC human beta 2 integrin alpha d subunit (see W23049). Its sequence  
 CC was deduced from a composite cDNA clone (see T79251). Homology  
 CC between the external domains of human and mouse alpha d is high,  
 CC but between cytoplasmic domains is low, suggesting C-terminal  
 CC functional differences. Recombinant alpha d polypeptides can be  
 CC expressed in transformed host cells and used to raise antibodies  
 CC or to assay for compounds that modulate alpha d activity. The  
 CC sequence of a murine alpha d polypeptide deduced from a full-length  
 CC cDNA clone is given in W23061.  
 SQ Sequence 1155 AA;

Query Match 69.6%; Score 48; DB 1; Length 1155;

Best Local Similarity 66.7%; Pred. No. 1.35e+02;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
Db 1 MVRGWILL 9  
| | | | |  
QY 1 MPRGVVVT 9

Search completed: Sat Apr 15 00:07:08 2000  
Job time : 42 secs.

**THIS PAGE BLANK (USPTO)**

\*\*\*\*\*  
W P S R L  
\*\*\*\*\* (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Sat Apr 15 00:07:26 2000; MagPar time 3.28 Seconds  
Tabular output not generated. 109.948 Million cell updates/sec

Title: >US-08-452-843-12  
Description: (1-9) from US08452843.ppe  
Perfect Score: 69  
Sequence: 1 MPRGVVVT 9

Scoring table: PAM 150  
Gap 15

Searched: 122810 seqs, 40068593 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: p1r62  
p1r1 2:p1r2 3:p1r3 4:p1r4

Statistics: Mean 23.655; Variance 29.310; scale 0.807

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES				Pred. No.	
Result No.	Score	Query Match	Description	ID	
1	55	79.7	laurate omega-hydroxy	A34160	8.09e-01
2	53	76.8	probable fusa2 protei	714 2 A70983	2.13e-00
3	52	75.4	ribosomal protein L5	181 1 RSKT5	3.42e-00
4	51	73.9	ribosomal protein L5	200 2 S77490	5.47e-00
5	51	73.9	cytochrome P450 4A7	511 2 B34160	5.47e-00
6	50	72.5	4-aminobutyrate trans	446 2 S72743	8.69e-00
7	49	71.0	collipase B precursor	96 1 XLHOB	1.37e-01
8	49	71.0	collipase A precursor	96 1 XLHOA	1.37e-01
9	49	71.0	Na+/taurocholate tran	362 2 A41601	1.37e-01
10	49	71.0	capsid protein - Sacc	697 2 S72353	1.37e-01
11	49	71.0	nitrate reductase (un	831 2 S50163	1.37e-01
12	49	71.0	hypothetical protein	1154 2 S43277	1.37e-01
13	49	71.0	hypothetical protein	1154 2 S43275	1.37e-01
14	49	71.0	genome polyprotein -	1512 2 S72354	1.37e-01
15	48	69.6	hypothetical 5K prote	54 2 JQ0639	2.15e-01
16	48	69.6	gerC2 protein - Helic	246 2 C84705	2.15e-01
17	48	69.6	hydroxysteroid sulfit	338 2 JF0196	2.15e-01
18	48	69.6	sensor kinase (EC 2.7	428 2 I39871	2.15e-01
19	48	69.6	vascular cell adhesio	538 2 Jc2457	2.15e-01
20	48	69.6	phosphoribosylaminoim	552 1 DEZPP	2.15e-01
21	48	69.6	phosphoribosylaminoim	552 2 S43322	2.15e-01
22	48	69.6	probable acyl-CoA syn	567 2 A70702	2.15e-01
23	48	69.6	phosphoribosylaminoim	571 1 DEBYP	2.15e-01

probable phosphoribos 2.15e-01  
pancreatic colipase - 3.34e-01  
colipase II precursor 3.34e-01  
colipase precursor - 3.34e-01  
immediate-early prote 3.34e-01  
acyltransferase homol 3.34e-01  
mitogenic factor, 25K 3.34e-01  
transposase (insertio 3.34e-01  
hypothetical protein 3.34e-01  
probable serine prote 3.34e-01  
H-transferring ATP s 3.34e-01  
ribN protein - Vibrio 3.34e-01  
myosin-V - chicken 3.34e-01  
myosin heavy chain, d 3.34e-01  
genome polyprotein - 3.34e-01  
genome polyprotein - 3.34e-01  
alpha-2-macroglobulin 3.34e-01  
retinoic acid recepto 5.17e-01  
galactokinase (EC 2.7 5.17e-01  
phosphoribosylaminoim 5.17e-01  
phosphoribosylaminoim 5.17e-01  
aldehyde oxidase (EC 5.17e-01

ALIGNMENTS

RESULT 1  
ENTRY  
TITLE  
ALTERNATE\_NAMES  
ORGANISM  
DATE  
ACCESSIONS  
REFERENCE  
#authors  
#journal  
#title  
#cross-references MUID:90094341  
#accession A34160  
#molecule\_type mRNA  
#residues 1-510 #label YOK  
#cross-references GB:M29531; NID:gl64986; PID:gl64987; GB:J05150  
REFERENCE  
#authors  
#journal  
#title  
#cross-references MUID:90254128  
#accession B34260  
#molecule\_type mRNA  
#residues 1-423,'VM',426-434,'R',436-475,'V',477-510 #label JOH  
#cross-references GB:M28656  
REFERENCE  
#authors  
#journal  
#title  
#cross-references MUID:90299866  
#accession PQ0047  
#molecule\_type protein  
#residues 5-24 #label KIK  
#experimental\_source liver  
#note amino-terminal sequence  
REFERENCE S23949

A34160 #type complete  
laurate omega-hydroxylase (EC 1.14.15.3) cytochrome P450 4A6  
- rabbit  
cytochrome P450ka-1; cytochrome P450LPGA omega 1  
#formal\_name Oryctolagus cuniculus #common\_name domestic  
rabbit  
31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change  
05-Mar-1999  
A34160; B34260; PQ0047; S23949  
A34160  
Yokotani, N.; Bernhardt, R.; Sogawa, K.; Kusunose, E.; Gotoh,  
O.; Kusunose, M.; Fujii-Kuriyama, Y.  
J. Biol. Chem. (1989) 264:21665-21669  
Two forms of omega-hydroxylase toward prostaglandin A and  
laurate. cDNA cloning and their expression.  
#cross-references MUID:90094341  
#accession A34160  
#molecule\_type mRNA  
#residues 1-510 #label YOK  
#cross-references GB:M29531; NID:gl64986; PID:gl64987; GB:J05150  
A34260  
Johnson, E.F.; Walker, D.L.; Griffin, K.J.; Clark, J.E.;  
Okita, R.T.; Muerthoff, A.S.; Masters, B.S.  
Biochemistry (1990) 29:873-879  
Cloning and expression of three rabbit kidney cDNAs encoding  
lauric acid omega-hydroxylases.  
#cross-references MUID:90254128  
#accession B34260  
#molecule\_type mRNA  
#residues 1-423,'VM',426-434,'R',436-475,'V',477-510 #label JOH  
#cross-references GB:M28656  
PQ0047  
Kikuta, Y.; Kusunose, E.; Okumoto, T.; Kubota, I.; Kusunose,  
M.  
J. Biochem. (1990) 107:280-286  
Purification and characterization of two forms of cytochrome  
P-450 with omega-hydroxylase activities toward  
prostaglandin A and fatty acids from rabbit liver  
microsomes.  
#cross-references MUID:90299866  
#accession PQ0047  
#molecule\_type protein  
#residues 5-24 #label KIK  
#experimental\_source liver  
#note amino-terminal sequence  
REFERENCE S23949

```

#authors Muerhoff, A.S.; Griffin, K.J.; Johnson, E.F.
#journal Arch. Biochem. Biophys. (1992) 296:66-72
#title Characterization of a rabbit gene encoding a
#cross-references MUID:92295782
#accession S23949
#status preliminary
#molecule_type DNA
#residues 1-26, 'C', 28-379, 'M', 381-423, 'YW', 426-434, 'R', 436-475,
#comment This enzyme catalyzes the omega-hydroxylation of prostaglandin A1
and A2, as well as the omega- and (omega-1)-hydroxylation of
fatty acid.
GENETICS
#gene CYP4A6
#classification #superfamily human cytochrome P450 CYP4B1; cytochrome P450
homology
#keywords chromoprotein; electron transfer; heme; iron; monooxygenase;
oxidoreductase; transmembrane protein
#feature 457
#binding site heme iron (Cys) (axial ligand) #status
predicted
#length 510 #molecular-weight 58129 #checksum 6263
Query Match 79.7%; Score 55; DB 2; Length 510;
Best Local Similarity 66.7%; Pred. NO. 8.09e-01;
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
Db 405 LKPGVIVTL 413
:|||||
Qy 1 MPRGVVVT 9

RESULT 2
ENTRY A70983 #type complete
TITLE probable fusa2 protein - Mycobacterium tuberculosis (strain
H37Rv)
ORGANISM #formal_name Mycobacterium tuberculosis
DATE 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change
16-Dec-1998
ACCESSIONS A70983
REFERENCE A70500
#authors Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher,
C.; Harris, D.; Gordon, S.V.; Eiglmeyer, K.; Gas, S.; Barry
III, C.E.; Tekala, F.; Badcock, K.; Basham, D.; Brown, D.;
Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.;
Fellwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;
Hornsby, L.; Jagers, K.; Krogh, A.; McLean, J.; Moule, S.;
Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.;
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.;
Skelton, S.; Squares, S.; Squares, R.; Sulston, J.E.;
Taylor, K.; Whitehead, S.; Barrell, B.G.
#journal Nature (1998) 393:537-544
#title Deciphering the biology of Mycobacterium tuberculosis from
the complete genome sequence.
#cross-references MUID:98295987
#accession A70983
#status preliminary; nucleic acid sequence not shown;
translation not shown
#molecule_type DNA
#residues 1-714 #label COL
#cross-references GB:296071; GB:AL123456; NID:g3242254; PID:e321091;
PID:g2181962
#experimental_source strain H37Rv
GENETICS
#gene fusa2
#classification #superfamily translation elongation factor Tu homology
#feature 24-151
#domain translation elongation factor Tu homology #label
EP1
#length 714 #molecular-weight 75630 #checksum 585
SUMMARY
Query Match 76.8%; Score 53; DB 2; Length 714;

```

---

```

Best Local Similarity 75.0%; Pred. NO. 2.13e+00;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
Db 141 MPRAVVIT 148
:|||||
Qy 1 MPRGVVVT 8

RESULT 3
ENTRY R5KT5 #type complete
TITLE ribosomal protein L5 - Cyanophora paradoxa cyanelle
ORGANISM #formal_name cyanelle Cyanophora paradoxa
DATE 31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change
05-Sep-1997
ACCESSIONS S07067; S12216
REFERENCE S07067; S12216
#authors Bryant, D.A.; Stirewalt, V.L.
#journal FEBS Lett. (1990) 259:273-280
#title The cyanelle genome of Cyanophora paradoxa encodes ribosomal
proteins not encoded by the chloroplast genomes of higher
plants.
#cross-references MUID:90092562
#accession S07067
#molecule_type DNA
#residues 1-181 #label BRY
#cross-references EMBL:X16548; NID:g11287; PID:g11288
REFERENCE S12211
#authors Michalowski, C.B.; Pfanzagl, B.; Loeffelhardt, W.; Bohnert,
H.J.
#journal Mol. Gen. Genet. (1990) 224:222-231
#title The cyanelle S10 spc ribosomal protein gene operon from
Cyanophora paradoxa.
#cross-references MUID:91117189
#accession S12216
#molecule_type DNA
#residues 1-145, 'G', 147-153, 'D', 165-181 #label MIC
#cross-references GB:M30487; NID:g336645; PID:g336651
GENETICS
#gene rpl5; rplE
#map_position 47-48
#genome cyanelle
#classification #superfamily Escherichia coli ribosomal protein L5
#keywords cyanelle; protein biosynthesis; ribosome
SUMMARY #length 181 #molecular-weight 20482 #checksum 584
Query Match 75.4%; Score 52; DB 1; Length 181;
Best Local Similarity 77.8%; Pred. NO. 3.42e+00;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
Db 84 MPIGVVVT 92
:|||||
Qy 1 MPRGVVVT 9

RESULT 4
ENTRY S77490 #type complete
TITLE ribosomal protein L5 - Synecocystis sp. (strain PCC 6803)
ALTERNATE_NAMES protein sll1808
ORGANISM #formal_name Synecocystis sp.
#variety PCC 6803
DATE 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change
21-Aug-1998
ACCESSIONS S77490; S34481
REFERENCE S74322
#authors Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.;
Nakamura, Y.; Miyajima, N.; Hirose, M.; Sugita, M.;
Sasamoto, S.; Kimura, T.; Hosouchi, T.; Matsuno, A.;
Muraki, A.; Nakazaki, N.; Naruo, K.; Okumura, S.; Shimo,
S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.;
Yasuda, M.; Tabata, S.
#journal DNA Res. (1996) 3:109-136
#title Sequence analysis of the genome of the unicellular
cyanobacterium Synecocystis sp. PCC6803. II. Sequence

```

```

determination of the entire genome and assignment of
potential protein-coding regions.
#cross-references MUID:97061201
#accession S77490
#status nucleic acid sequence not shown; translation not shown
#molecule_type DNA
#residues 1-200 #label KAN
#cross-references EMBL:D90905; GB:AB001339; NID:g1652360; PID:d1018070;
PID:g1652415
##note the nucleotide sequence was submitted to the EMBL Data
Library, June 1996
REFERENCE S34477
#authors Berger, S.; Ellersiek, U.; Kinzelt, D.; Steilmueller, K.
#journal FEBS Lett. (1993) 326:246-250
#title Immunoprecipitation of a subcomplex of the NAD
(P)H-plastoquinone-oxidoreductase from the cyanobacterium
Synechocystis sp. PCC6803.
#cross-references MUID:93314795
#accession S34481
#molecule_type protein
#residues 22-30;32-33 #label BER
GENETICS
#gene rpl5
#start_codon GTG
CLASSIFICATION #superfamily Escherichia coli ribosomal protein L5
KEYWORDS protein biosynthesis; ribosome
SUMMARY #length 200 #molecular-weight 22508 #checksum 5351
Query Match 73.9%; Score 51; DB 2; Length 200;
Best Local Similarity 77.8%; Pred. No. 5.47e+00;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
Db 104 MPVGVVMTL 112
||| |||
QY 1 MPRGVVMTL 9
RESULT 5
ENTRY B34160 #type complete
TITLE cytochrome P450 4A7 - rabbit
ALTERNATE_NAMES cytochrome P450ka-2; cytochrome P450kc
CONTAINS oxidoreductase (EC 1.-.-)
ORGANISM #formal_name Oryctolagus cuniculus #common_name domestic
rabbit
DATE 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change
05-Mar-1999
ACCESSIONS B34160; C34260; JN0090
REFERENCE B34160
#authors Yokotani, N.; Bernhardt, R.; Sogawa, K.; Kusunose, E.; Gotoh,
O.; Kusunose, M.; Fujii-Kuriyama, Y.
#journal J. Biol. Chem. (1989) 264:21665-21669
#title Two forms of omega-hydroxylase toward prostaglandin A and
laurate. cDNA cloning and their expression.
#cross-references MUID:90094341
#accession B34160
#molecule_type mRNA
#residues 1-511 #label YOK
#cross-references GB:M29530; NID:g164984; PID:g164985; GB:J05150
REFERENCE A34260
#authors Johnson, E.F.; Walker, D.L.; Griffin, K.J.; Clark, J.E.;
Okita, R.T.; Muerhoff, A.S.; Masters, B.S.
#journal Biochemistry (1990) 29:873-879
#title Cloning and expression of three rabbit kidney cDNAs encoding
lauric acid omega-hydroxylases.
#cross-references MUID:90254128
#accession C34260
#molecule_type mRNA
#residues 1-98,'C',100-149,'F',151-391,'SK',394-476,'V',478-511
##label JOH
#cross-references GB:M28657; NID:g164978; PID:g164979
REFERENCE JN0089
#authors Yoshimura, R.; Kusunose, E.; Yokotani, N.; Yamamoto, S.;
Kubota, I.; Kusunose, M.

```

```

#journal J. Biochem. (1990) 108:544-548
#title Purification and characterization of two forms of fatty acid
omega-hydroxylase cytochrome P-450 from rabbit kidney
cortex microsomes.
#cross-references MUID:91154157
#accession JN0090
#molecule_type protein
#residues 5-7,'X',9-15,'X',17-22,'X',24 #label YOS
#experimental_source kidney
COMMENT This protein catalyzes the omega- and (omega-1)-hydroxylation of
fatty acids.
CLASSIFICATION #superfamily human cytochrome P450 CYP4B1; cytochrome P450
homology
KEYWORDS chromoprotein; electron transfer; heme; iron; monooxygenase;
oxidoreductase; transmembrane protein
FEATURE
458 #binding_site heme iron (cys) (axial ligand) #status
predicted
SUMMARY #length 511 #molecular-weight 58337 #checksum 4545
Query Match 73.9%; Score 51; DB 2; Length 511;
Best Local Similarity 44.4%; Pred. No. 5.47e+00;
Matches 4; Conservative 5; Mismatches 0; Indels 0; Gaps 0;
Db 406 LPKGIITL 414
:||:|
QY 1 MPRGVVMTL 9
RESULT 6
ENTRY S72743 #type complete
TITLE 4-aminobutyrate transaminase (EC 2.6.1.19) gabT -
Mycobacterium leprae
ALTERNATE_NAMES 4-aminobutyrate aminotransferase gabT; B1177_F2_67 protein
ORGANISM #formal_name Mycobacterium leprae cosmid B1177.
DATE 19-Mar-1997 #sequence_revision 25-Apr-1997 #text_change
09-Sep-1997
ACCESSIONS S72743
REFERENCE S72693
#authors Smith, D.R.; Robison, K.
#submission Submitted to the EMBL Data Library, November 1993
#description Mycobacterium leprae cosmid B1177.
#accession S72743
#status preliminary
#molecule_type DNA
#residues 1-446 #label SMI
#cross-references EMBL:U00011; NID:g466807; PID:g466832
GENETICS
#gene gabT
#start_codon GTG
KEYWORDS aminotransferase
SUMMARY #length 446 #molecular-weight 47215 #checksum 5414
Query Match 72.5%; Score 50; DB 2; Length 446;
Best Local Similarity 77.8%; Pred. No. 8.69e+00;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
Db 30 VPRGVVMTL 38
:|||
QY 1 MPRGVVMTL 9
RESULT 7
ENTRY XLHOB #type complete
TITLE Colipase B precursor - horse
ALTERNATE_NAMES procolipase B
ORGANISM #formal_name Equus caballus #common_name domestic horse
DATE 14-Nov-1983 #sequence_revision 14-Nov-1983 #text_change
26-Apr-1996
ACCESSIONS A03165; B90220
REFERENCE A90637
#authors Bonicel, J.; Couchoud, P.; Foglizzo, E.; Desnuelle, P.;
Chapus, C.

```

```
#journal Biochim. Biophys. Acta (1981) 669:39-45
#title Amino acid sequence of horse colipase B.
#cross-references MUID:82046794
#accession A03165
#molecule_type protein
#residues 1-96 #label BON
REFERENCE
#authors Julien, R.; Bechis, G.; Gregoire, J.; Rathelot, J.; Rochat, H.; Sarda, L.
#journal Biochem. Biophys. Res. Commun. (1980) 95:1245-1252
#title Evidence for the existence of two isocolipases in horse pancreas.
#cross-references MUID:81021166
#accession B90220
#molecule_type protein
#residues 1-21,'E',23-28,'T',30-55 #label JUL
COMMENT Colipase, a cofactor of triacylglycerol lipase (EC 3.1.1.3), forms a 1:1 stoichiometric complex with it, enabling it to hydrolyze its substrate at the lipid-water interface. Without colipase the enzyme is washed off by bile salts, which are known to have an inhibitory effect on the lipase.
CLASSIFICATION #superfamily colipase
KEYWORDS lipid digestion; lipid hydrolysis; pancreas
FEATURE
1-5 #domain propeptide #status experimental #label PRO\
6-96 #product colipase A #status experimental #label MAT\
17-87,23-39,27-63, #disulfide_bonds #status predicted\
28-61,49-69 #binding_site micellar substrate (Tyr, Tyr, Tyr, Tyr)
52,55,58,59 #status predicted
SUMMARY #length 96 #molecular-weight 10491 #checksum 3464
Query Match 71.0%; Score 49; DB 1; Length 96;
Best Local Similarity 62.5%; Pred. No. 1.37e+01;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Db 4 PRGVIINL 11
||||:|
QY 2 PRGVVWTL 9

RESULT 8
ENTRY XLHOA #type complete
TITLE colipase A precursor - horse
ALTERNATE_NAMES procolipase A
ORGANISM #formal_name Equus caballus #common_name domestic horse
DATE 14-Nov-1993 #sequence_revision 04-Dec-1986 #text_change
ACCESSIONS A03164; A91119; A90220
REFERENCE A90652
#authors Sternby, B.; Engstrom, A.; Hellman, U.; Vihter, A.M.; Sternby, N.H.; Borgstrom, B.
#journal Biochim. Biophys. Acta (1984) 784:75-80
#title The primary sequence of human pancreatic colipase.
#cross-references MUID:84104937
#accession A03164
#molecule_type protein
#residues 1-96 #label SPT
#note residues 56-59 were positioned by homology; no overlap was obtained for 65-66
REFERENCE
#authors A91119
#journal Eur. J. Biochem. (1982) 123:347-354
#title Pancreatic colipase: crystallographic and biochemical aspects.
#cross-references MUID:82186702
#accession A91119
#molecule_type protein
#residues 1-88,'N',90-91,'K',93 #label PTE
REFERENCE
#authors Julien, R.; Bechis, G.; Gregoire, J.; Rathelot, J.; Rochat, H.; Sarda, L.
```

```
#journal Biochem. Biophys. Res. Commun. (1980) 95:1245-1252
#title Evidence for the existence of two isocolipases in horse pancreas.
#cross-references MUID:81021166
#accession A90220
#molecule_type protein
#residues 1-21,'Q',23-55 #label JUL
COMMENT Colipase, a cofactor of triacylglycerol lipase (EC 3.1.1.3), forms a 1:1 stoichiometric complex with it, enabling it to hydrolyze its substrate at the lipid-water interface. Without colipase the enzyme is washed off by bile salts, which are known to have an inhibitory effect on the lipase.
CLASSIFICATION #superfamily colipase
KEYWORDS lipid digestion; lipid hydrolysis; pancreas
FEATURE
1-5 #domain propeptide #status experimental #label PRO\
6-96 #product colipase A #status experimental #label MAT\
17-87,23-39,27-63, #disulfide_bonds #status predicted\
28-61,49-69 #binding_site micellar substrate (Tyr, Tyr, Tyr, Tyr)
52,55,58,59 #status predicted
SUMMARY #length 96 #molecular-weight 10488 #checksum 2704
Query Match 71.0%; Score 49; DB 1; Length 96;
Best Local Similarity 62.5%; Pred. No. 1.37e+01;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Db 4 PRGVIINL 11
||||:|
QY 2 PRGVVWTL 9

RESULT 9
ENTRY A41601 #type complete
TITLE Na+/taurocholate transport protein - rat
ORGANISM #formal_name Rattus norvegicus #common_name Norway rat
DATE 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change
ACCESSIONS A41601
REFERENCE A41601
#authors Hagenbuch, B.; Stieger, B.; Foguet, M.; Luebbert, H.; Maier, P.J.
#journal Proc. Natl. Acad. Sci. U.S.A. (1991) 88:10629-10633
#title Functional expression cloning and characterization of the hepatocyte Na(+)/bile acid cotransport system.
#cross-references MUID:92073340
#accession A41601
#status preliminary
#molecule_type mRNA
#residues 1-362 #label HAG
#cross-references GB:M77429
KEYWORDS transmembrane protein
SUMMARY #length 362 #molecular-weight 39295 #checksum 8711
Query Match 71.0%; Score 49; DB 2; Length 362;
Best Local Similarity 55.6%; Pred. No. 1.37e+01;
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
Db 57 KPGKVVVAL 65
||||:|
QY 1 MPKGVVWTL 9

RESULT 10
ENTRY S72353 #type complete
TITLE capsid protein - Saccharomyces cerevisiae virus La
ORGANISM #formal_name Saccharomyces cerevisiae virus La, Scv-La
DATE 04-May-1998 #sequence_revision 22-May-1998 #text_change
ACCESSIONS S72353; S50258
REFERENCE S72353
#authors Park, C.M.; Lopinski, J.D.; Masuda, J.; Tzeng, T.H.; Bruenn, J.A.
```



```
#journal      Virology (1996) 216:451-454
#title       A second double-stranded RNA virus from yeast.
#cross-references MUID:96182949
#accession   S72353
#molecule_type genomic RNA
#residues_type 1-697 #label PAR
#cross-references EMBL:U01060; NID:g595249; PID:g595250

GENETICS
#gene        capsid protein
#keywords    #length 697 #molecular-weight 78315 #checksum 3673
SUMMARY
Query Match      71.0%; Score 49; DB 2; Length 697;
Best Local Similarity 55.6%; Pred. No. 1.37e+01;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 393 MNRGIIVDL 401
QY 1 MPRGVVVTL 9

RESULT 11
ENTRY #S0163 #type complete
TITLE nitrate reductase (unclassified) (EC 1.-.-.-) large chain
        precursor, periplasmic - Thiosphaera pantotropha
ORGANISM #formal_name Thiosphaera pantotropha
DATE 16-Feb-1995 #sequence_revision 26-Jul-1996 #text_change
        21-Aug-1998
ACCESSIONS S50163; S56135; S56128
REFERENCE Barks, B.C.; Richardson, D.J.; Reilly, A.; Willis, A.C.;
        Ferguson, S.J.
#submission submitted to the EMBL Data Library, August 1994
#description The periplasmic nitrate reductase operon of Thiosphaera
        pantotropha.
#accession S50163
#molecule_type DNA
#residues 1-831 #label BER
#cross-references EMBL:Z36773; NID:g600089; PID:g600093
REFERENCE S56128
#authors Barks, B.C.; Richardson, D.J.; Reilly, A.; Willis, A.C.;
        Ferguson, S.J.
#journal Biochem. J. (1995) 309:983-992
#title The napEDABC gene cluster encoding the periplasmic nitrate
        reductase system of Thiosphaera pantotropha.
#accession S56135
#status nucleic acid sequence not shown
#molecule_type DNA
#residues 1-31; 42-89; 154-203 #label BEW
#cross-references EMBL:Z36773
#accession S56128
#molecule_type protein
#residues 114-127; 'D', 129-130; 139-157; 317-339; 451-469; 591-600;
        642-657; 675-694; 699-715 #label BEF

GENETICS
#gene napA
#classification #superfamily formate dehydrogenase
#keywords blocked amino end; oxidoreductase
FEATURE
1-31 #domain signal sequence #status predicted #label SIG\
32-831 #product nitrate reductase large chain #status predicted
        #label MAT
SUMMARY #length 831 #molecular-weight 92617 #checksum 1537
Query Match      71.0%; Score 49; DB 2; Length 831;
Best Local Similarity 100.0%; Pred. No. 1.37e+01;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 788 MPRGVV 793
QY 1 MPRGVV 6
```

```
RESULT 12
ENTRY #S43277 #type complete
TITLE hypothetical protein 2 - Neurospora crassa retrotransposon
        Tad3-2
ORGANISM #formal_name Neurospora crassa
DATE 19-Mar-1997 #sequence_revision 25-Apr-1997 #text_change
        09-Sep-1997
ACCESSIONS S43277
REFERENCE Cambareri, E.B.; Helber, J.; Kinsey, J.A.
        Mol. Gen. Genet. (1994) 242:658-665
#title Tad1-1, an active LINE-like element of Neurospora crassa.
#cross-references MUID:94203179
#accession S43277
#status preliminary; nucleic acid sequence not shown;
        translation not shown
#molecule_type DNA
#residues 1-1154 #label CAM
#cross-references EMBL:L25863; NID:g409762; PID:g409764
#note the nucleotide sequence was submitted to the EMBL Data
        Library, November 1993
SUMMARY #length 1154 #molecular-weight 130470 #checksum 5717
Query Match      71.0%; Score 49; DB 2; Length 1154;
Best Local Similarity 55.6%; Pred. No. 1.37e+01;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 828 KPRGIVIGL 836
QY 1 MPRGVVVTL 9

RESULT 13
ENTRY #S43275 #type complete
TITLE hypothetical protein 2 - Neurospora crassa retrotransposon
        Tad1-1
ORGANISM #formal_name Neurospora crassa
DATE 19-Mar-1997 #sequence_revision 25-Apr-1997 #text_change
        09-Sep-1997
ACCESSIONS S43275
REFERENCE Cambareri, E.B.; Helber, J.; Kinsey, J.A.
        Mol. Gen. Genet. (1994) 242:658-665
#title Tad1-1, an active LINE-like element of Neurospora crassa.
#cross-references MUID:94203179
#accession S43275
#molecule_type DNA
#residues 1-1154 #label CAM
#cross-references EMBL:L25862; NID:g409759; PID:g409761
GENETICS
#mobile_element retrotransposon Tad1-1
SUMMARY #length 1154 #molecular-weight 130398 #checksum 5771
Query Match      71.0%; Score 49; DB 2; Length 1154;
Best Local Similarity 55.6%; Pred. No. 1.37e+01;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 828 KPRGIVIGL 836
QY 1 MPRGVVVTL 9

RESULT 14
ENTRY #S72354 #type complete
TITLE genome polyprotein - Saccharomyces cerevisiae virus L-A
ALTERNATE_NAMES cap-pol fusion protein
CONTAINS RNA-directed RNA polymerase (EC 2.7.7.48)
ORGANISM #formal_name Saccharomyces cerevisiae virus L-A, SCV-L-A
DATE 04-May-1998 #sequence_revision 22-May-1998 #text_change
        01-Feb-1999
ACCESSIONS S72354; S12851; S41483; S50259; S14223
REFERENCE S72353
#authors Park, C.M.; Lopinski, J.D.; Masuda, J.; Tzeng, T.H.; Bruenn,
```

J.A.  
#journal Virology (1996) 216:451-454  
#title A second double-stranded RNA virus from yeast.  
#cross-references MUID:96182949  
#accession S72354  
##molecule\_type genomic RNA  
##residues 1-1512 #label PAR  
##cross-references EMBL:U01060; NID:9595249; PID:9595251  
##note biosynthesis of this protein involves a -1 frameshift in the codon for residue 649

REFERENCE  
#authors Bruenn, J.A.  
#journal Nucleic Acids Res. (1991) 19:217-226  
#title Relationships among the positive strand and double-strand RNA viruses as viewed through their RNA-dependent RNA polymerases.  
#cross-references MUID:91195040  
#accession S12851  
##status nucleic acid sequence not shown  
##molecule\_type genomic RNA  
##residues 1271-1512 #label BRW  
##cross-references EMBL:X54405; NID:g61953; PID:g61954  
##note the sequence of residue 1512 and the corresponding nucleic acid sequence are not shown

REFERENCE  
#authors Bruenn, J.A.  
#journal Nucleic Acids Res. (1993) 21:5667-5669  
#title A closely related group of RNA-dependent RNA polymerases from double-stranded RNA viruses.  
#cross-references MUID:94111988  
#accession S41483  
##molecule\_type genomic RNA  
##residues 956-959; 1018-1027; 1079-1084; 1136-1146; 1199-1218; 1232-1239; 1273-1278; 1290-1293 #label BRU

GENETICS  
#gene cap/pol  
CLASSIFICATION #superfamily Saccharomyces cerevisiae virus La RNA-directed RNA polymerase  
KEYWORDS nucleotidyltransferase; translational frameshift  
SUMMARY #length 1512 #molecular-weight 171448 #checksum 7155

Query Match 71.08; Score 49; DB 2; Length 1512;  
Best Local Similarity 55.6%; Pred. No. 1.37e+01;  
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 393 MPRGIIVDL 401  
| | | | |  
QY 1 MPRGVVTL 9

RESULT 15  
ENTRY JQ0639 #type complete  
TITLE hypothetical 5K protein (hisb5 5' region) - Streptomyces coelicolor  
ORGANISM #formal\_name Streptomyces coelicolor  
DATE 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 09-Sep-1997  
ACCESSIONS JQ0639  
REFERENCE JQ0637  
#authors Limauro, D.; Avitabile, A.; Cappellano, C.; Puglia, A.M.; Bruni, C.B.  
#journal Gene (1990) 90:31-41  
#title Cloning and characterization of the histidine biosynthetic gene cluster of Streptomyces coelicolor A3(2).  
#cross-references MUID:90337345  
#accession JQ0639  
##molecule\_type DNA  
##residues 1-54 #label LIM  
##cross-references GB:M31628; NID:g153295; PID:g153299  
##experimental\_source strain A3[2]  
SUMMARY #length 54 #molecular-weight 5686 #checksum 3200

Query Match 69.6%; Score 48; DB 2; Length 54;

Best Local Similarity 55.6%; Pred. No. 2.15e+01;  
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 27 MPKGLIVLL 35  
| | | | |  
QY 1 MPRGVVTL 9

Search completed: Sat Apr 15 00:07:35 2000  
Job time : 9 secs.

\*\*\*\*\*  
M P S R L H  
\*\*\*\*\*  
(TM)

Release 3.1A John F. Collins, BioComputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Sat Apr 15 00:08:58 2000; MasPar time 11.53 Seconds  
54.130 Million cell updates/sec

Tabular output not generated.

Title: >US-08-452-843-12  
Description: (1-9) from US08452843.pep  
Perfect Score: 69  
Sequence: 1 MPRGVVVT 9

Scoring table: PAM 150  
Gap 15

Searched: 225878 seqs, 69334122 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: sptrenb112  
1:sp-archaea 2:sp-bacteria 3:sp-fungi 4:sp-human  
5:sp-invertebrate 6:sp-mammal 7:sp-mhc 8:sp-organelle  
9:sp-phage 10:sp-plant 11:sp-rodent 12:sp-unclassified  
13:sp-vertebrate 14:sp-virus

Statistics: Mean 23.627; Variance 24.501; scale 0.964

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES				
Result No.	Score	Query Match	Description	Pred. No.
1	53	76.8	FUS2.	7.04e-01
2	50	72.5	125A2. LONG HYPOTHETICA	3.72e+00
3	49	71.0	149A. LONG HYPOTHETICA	6.38e+00
4	49	71.0	172.14 Q9YND6 STRUCTURAL PROTEIN VP2	6.38e+00
5	49	71.0	1-AMINOCYCLOPROPANE-1-SEPL.	6.38e+00
6	49	71.0	351.2 Q52149 SEPL.	6.38e+00
7	49	71.0	351.2 Q69411 SEPL.	6.38e+00
8	49	71.0	351.2 Q86194 SEPL. PROTEIN.	6.38e+00
9	49	71.0	351.2 Q9W16 PMBA-RELATED PROTEIN.	6.38e+00
10	49	71.0	470.5 Q01487 COSMID C13F10.	6.38e+00
11	49	71.0	648.2 Q09875 CELL DIVISION PROTEIN	6.38e+00
12	49	71.0	687.14 Q87026 CAPSID.	6.38e+00
13	49	71.0	831.2 Q88111 PERIPLASMIC NITRATE RE	6.38e+00
14	49	71.0	1154.3 Q01375 HYPOTHETICAL 130.4 KD	6.38e+00
15	49	71.0	1154.3 Q01379 HYPOTHETICAL 130.5 KD	6.38e+00
16	49	71.0	2638.2 Q03914 DAPTOMYCIN BIOSYNTHETI	6.38e+00
17	49	71.0	3097.5 Q61143 NAD(P)H-DEPENDENT GLUT	6.38e+00
18	48	69.6	206.10 Q41860 TRANSPOSABLE ELEMENT M	1.08e+01
19	48	69.6	246.2 Q26017 GERC2 PROTEIN (GERC2).	1.08e+01
20	48	69.6	337.4 Q9Y271 CYSTEINYL LEUKOTRIENE	1.08e+01

21	48	69.6	338.11	Q35400	HYDROXYSTEROID SULFOTR	1.08e+01
22	48	69.6	350.4	O75814	HYDROXYSTEROID SULFOTR	1.08e+01
23	48	69.6	350.4	O00204	HYDROXYSTEROID SULFOTR	1.08e+01
24	48	69.6	365.4	O00205	HYDROXYSTEROID SULFOTR	1.08e+01
25	48	69.6	503.2	O65844	PUTATIVE CATIONIC AMIN	1.08e+01
26	48	69.6	526.2	Q53679	SIMILARITY TO GALACTOS	1.08e+01
27	48	69.6	538.6	Q29123	VASCULAR CELL ADHESION	1.08e+01
28	48	69.6	538.6	Q28939	VASCULAR CELL ADHESION	1.08e+01
29	48	69.6	550.5	O17145	GAMMA-AMINOBUTYRIC ACI	1.08e+01
30	48	69.6	562.2	P71605	HYPOTHETICAL 61.0 KD P	1.08e+01
31	48	69.6	570.3	O74197	PHOSPHORIBOSYLAMINOIM	1.08e+01
32	48	69.6	645.10	O80937	PUTATIVE PHOSPHORIBOSY	1.08e+01
33	48	69.6	915.5	O63328	R05H10.6 PROTEIN.	1.08e+01
34	47	68.1	404.1	O59488	404AA LONG HYPOTHETICA	1.83e+01
35	47	68.1	2183.14	Q9W984	RNA POLYMERASE.	1.83e+01
36	47	68.1	2183.14	Q9WPY7	RNA POLYMERASE.	1.83e+01
37	47	68.1	2183.14	Q9WPY8	RNA POLYMERASE.	1.83e+01
38	47	68.1	2183.14	Q9WPY9	RNA POLYMERASE.	1.83e+01
39	47	68.1	2183.14	Q9WMB3	LARGE POLYMERASE.	1.83e+01
40	47	68.1	2183.14	Q9WPY6	RNA POLYMERASE.	1.83e+01
41	47	68.1	2183.14	Q9WPY3	RNA POLYMERASE.	1.83e+01
42	47	68.1	2183.14	Q9WP20	RNA POLYMERASE.	1.83e+01
43	47	68.1	2183.14	Q9WP21	RNA POLYMERASE.	1.83e+01
44	47	68.1	2183.14	Q83626	LARGE POLYMERASE.	1.83e+01
45	47	68.1	2183.14	P90461	RNA DEPENDENT RNA POLY	1.83e+01

ALIGNMENTS

RESULT 1  
ID C07170 PRELIMINARY; PRT; 714 AA.  
AC C07170;  
DT 01-JUL-1997 (Tremblrel. 04, Created)  
DT 01-JUL-1997 (Tremblrel. 04, Last sequence update)  
DT 01-NOV-1999 (Tremblrel. 12, Last annotation update)  
DE FUS2.  
GN FUS2.  
OS Mycobacterium tuberculosis.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-H37RV;  
RA OLIVER K., HARRIS D.;  
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-H37RV;  
RA PARKHILL J., BARRELL B.G., RAJANDREAM M.A.;  
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-H37RV;  
RX MEDLINE; 96161548.  
RA PHILIPP W.J., POULET S., EIGLMEIER K., PASCOPELLA L.,  
RA BALASUBRAMANIAN V., HEYM B., BERGH S., BLOOM B.R., JACOBS W.R. JR.,  
RA COLE S.T.;  
RT "An integrated map of the genome of the tubercle bacillus,  
Mycobacterium tuberculosis H37Rv, and comparison with Mycobacterium  
leprae.";  
RL Proc. Natl. Acad. Sci. U.S.A. 93:3132-3137(1996).  
DR EMBL; Z96071; CAB09448.1; -;  
DR HSSP; P13551; 2EPG.  
DR PFAM; PF00679; EFG.C.1.  
DR PFAM; PF00009; GTP\_EFTU; 1.  
SQ SEQUENCE 714 AA; 75630 MW; D3E5E4E8 CRC32;  
Query Match 76.8%; Score 53; DB 2; Length 714;  
Best Local Similarity 75.0%; Pred. No. 7.04e-01;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
Db 141 MPRVVIT 148  
|||||:

```
QY 1 MPRGVVVT 8

RESULT 2
ID QYBF7 PRELIMINARY; PRT; 125 AA.
AC QYBF7;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE 125AA LONG HYPOTHETICAL PROTEIN.
GN APE1640.
OS Aeropyrum pernix.
OC Archaea; Crenarchaeota; Aeropyrum.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-KJ;
RX MEDLINE; 99310339.
RA KAWARABAYASI Y., HINO Y., HORIKAWA H., YAMAZAKI S., HAIKAWA Y.,
RA JIN-NO K., TAKAHASHI M., SEKINE M., BABA S., ANKAI A., KOSUGI H.,
RA HOSOIYAMA A., FUKUI S., NAGAI Y., NISHIJIMA K., NAKAZAWA H.,
RA TAKAMIYA M., MASUDA S., FUNAHASHI T., TANAKA T., KUDOH Y.,
RA YAMAZAKI J., KUSHIDA N., OGUCHI A., AOKI K., KUBOTA K., NAKAMURA Y.,
RA NOMURA N., SAKO Y., KIKUCHI H.;
RT "Complete genome sequence of an aerobic hyper-thermophilic
RT crenarchaeon, Aeropyrum pernix K1.";
RL DNA Res. 6:83-101(1999).
DR EMBL; AP000062; BAA80641.1; -.
SQ SEQUENCE 125 AA; 13587 MW; 08E4F208 CRC32;

Query Match 72.5%; Score 50; DB 1; Length 125;
Best Local Similarity 75.0%; Pred. No. 3.72e+00;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 94 MPRGVVVA 101
|||||
QY 1 MPRGVVVT 8

RESULT 3
ID QY9U6 PRELIMINARY; PRT; 149 AA.
AC QY9U6;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE 149AA LONG HYPOTHETICAL PROTEIN.
GN APE2193.
OS Aeropyrum pernix.
OC Archaea; Crenarchaeota; Aeropyrum.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-KJ;
RX MEDLINE; 99310339.
RA KAWARABAYASI Y., HINO Y., HORIKAWA H., YAMAZAKI S., HAIKAWA Y.,
RA JIN-NO K., TAKAHASHI M., SEKINE M., BABA S., ANKAI A., KOSUGI H.,
RA HOSOIYAMA A., FUKUI S., NAGAI Y., NISHIJIMA K., NAKAZAWA H.,
RA TAKAMIYA M., MASUDA S., FUNAHASHI T., TANAKA T., KUDOH Y.,
RA YAMAZAKI J., KUSHIDA N., OGUCHI A., AOKI K., KUBOTA K., NAKAMURA Y.,
RA NOMURA N., SAKO Y., KIKUCHI H.;
RT "Complete genome sequence of an aerobic hyper-thermophilic
RT crenarchaeon, Aeropyrum pernix K1.";
RL DNA Res. 6:83-101(1999).
DR EMBL; AP000063; BAA81204.1; -.
SQ SEQUENCE 149 AA; 17298 MW; 41ED2510 CRC32;

Query Match 71.0%; Score 49; DB 1; Length 149;
Best Local Similarity 100.0%; Pred. No. 6.38e+00;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 6 MPRGVV 11
|||||
QY 1 MPRGVV 6

RESULT 4
ID QYND6 PRELIMINARY; PRT; 172 AA.
AC QYND6;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE STRUCTURAL PROTEIN VP2 (FRAGMENT).
GN VP2.
OS Infectious bursal disease virus.
OC Viruses; dsRNA viruses; Birnaviridae; Avibirnavirus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-LABORATORY-SELECTED IBV STRAIN EM3;
RA ETERRADOSSI N., ARNAULD C., TOQUIN D., RIVALLAN G.;
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; Y14963; CAA75185.1; -.
FT NON_TER 1
FT NON_TER 172
SQ SEQUENCE 172 AA; 18048 MW; 077B1E7D CRC32;

Query Match 71.0%; Score 49; DB 14; Length 172;
Best Local Similarity 75.0%; Pred. No. 6.38e+00;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 40 PRGVTTTL 47
|||||
QY 2 PRGVVVTL 9

RESULT 5
ID P93459 PRELIMINARY; PRT; 314 AA.
AC P93459;
DT 01-MAY-1997 (TrEMBLrel. 03, Created)
DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE 1-AMINOCYCLOPROPANE-1-CARBOXYLIC ACID OXIDASE.
GN ACO.
OS Rumex palustris.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Caryophyllales; Caryophyllaceae; Polygonaceae; Rumex.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-LEAF;
RA VRIEZEN H.W., HULZINK R.;
RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; Y10034; CAA71140.1; -.
DR MENDEL; 12274; RumpAco; 12274.
DR PFAM; PF00671; Fe_Asc_Oxidored; 1.
SQ SEQUENCE 314 AA; 35665 MW; 5E318BA7 CRC32;

Query Match 71.0%; Score 49; DB 10; Length 314;
Best Local Similarity 55.6%; Pred. No. 6.38e+00;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 208 MPH5IVVNL 216
|||||
QY 1 MPRGVVVTL 9

RESULT 6
ID O52149 PRELIMINARY; PRT; 351 AA.
AC O52149;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-JUN-1998 (TrEMBLrel. 06, Last annotation update)
DE SEPL.
GN SEPL.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
RN [1]
RP SEQUENCE FROM N.A.
```

RA ELLIOTT S.J., WAINWRIGHT L.A., MCDANIEL T.K., JARVIS K.G., DENG Y.K.,  
RA LAI L.C., MCNAMARA B.P., DONNENBERG M.S., KAPER J.B.;  
RL MOL. Microbiol. 0:0-0(1998).  
DR EMBL: AF022236; AAC38393.1; -.  
SQ SEQUENCE 351 AA; 39987 MW; 1B238DB3 CRC32;

Query Match 71.0%; Score 49; DB 2; Length 351;  
Best Local Similarity 55.6%; Pred. No. 6.38e+00;  
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 124 MPKGEIVAL 132  
||:|:|:|  
Qy 1 MPRGVVVT 9

RESULT 7  
ID O69411 PRELIMINARY; PRT; 351 AA.  
AC O69411;  
DT 01-AUG-1998 (TReMBLrel. 07, Created)  
DT 01-AUG-1998 (TReMBLrel. 07, Last sequence update)  
DT 01-NOV-1998 (TReMBLrel. 08, Last annotation update)  
DE SEPL.  
GN SEPL.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Escherichia.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-EHEC EDL933;  
RA KRESSE A.U., EBEL F., DEIBEL C., CHAKRABORTY T., GUZMAN C.A.;  
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-EDL933;  
RX MEDLINE; 98339885.  
RA PERNA N.T., MAYHEW G.F., POSFAI G., ELLIOTT S., DONNENBERG M.S.,  
RA KAPER J.B., BLATTNER F.R.;  
RT "Molecular evolution of a pathogenicity island from enterohemorrhagic  
RT Escherichia coli O157:H7."  
RL Infect. Immun. 66:3810-3817(1998).  
DR EMBL: Y13068; CAA73505.1; -.  
DR EMBL: AF071034; AAC31502.1; -.  
SQ SEQUENCE 351 AA; 39951 MW; 734E2361 CRC32;

Query Match 71.0%; Score 49; DB 2; Length 351;  
Best Local Similarity 55.6%; Pred. No. 6.38e+00;  
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 124 MPKGEIVAL 132  
||:|:|:|  
Qy 1 MPRGVVVT 9

RESULT 8  
ID O86194 PRELIMINARY; PRT; 351 AA.  
AC O86194;  
DT 01-NOV-1998 (TReMBLrel. 08, Created)  
DT 01-NOV-1998 (TReMBLrel. 08, Last sequence update)  
DT 01-NOV-1998 (TReMBLrel. 08, Last annotation update)  
DE SEPL PROTEIN.  
GN SEPL.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Escherichia.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-413/89-1;  
RX MEDLINE; 97045129.  
RA EBEL F., DEIBEL C., KRESSE A.U., GUZMAN C., CHAKRABORTY T.;  
RT "Temperature- and medium-dependent secretion of proteins by Shiga  
RT toxin-producing Escherichia coli."  
RL Infect. Immun. 64:4472-4479(1996).  
RN [2]

RP SEQUENCE FROM N.A.  
RC STRAIN-413/89-1;  
RX MEDLINE; 98389647.  
RA KRESSE A.U., SCHULZE K., DEIBEL C., EBEL F., ROHDE M., CHAKRABORTY T.,  
RA GUZMAN C.A.;  
RT "Pas, a novel protein required for protein secretion and attaching and  
RT effacing activities of enterohemorrhagic Escherichia coli."  
RL J. Bacteriol. 180:4370-4379(1998).  
DR EMBL: Y13859; CAA74171.1; -.  
SQ SEQUENCE 351 AA; 39974 MW; 8D815D78 CRC32;

Query Match 71.0%; Score 49; DB 2; Length 351;  
Best Local Similarity 55.6%; Pred. No. 6.38e+00;  
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 124 MPKGEIVAL 132  
||:|:|:|  
Qy 1 MPRGVVVT 9

RESULT 9  
ID O9WZ16 PRELIMINARY; PRT; 435 AA.  
AC O9WZ16;  
DT 01-NOV-1999 (TReMBLrel. 12, Created)  
DT 01-NOV-1999 (TReMBLrel. 12, Last sequence update)  
DT 01-NOV-1999 (TReMBLrel. 12, Last annotation update)  
DE PMBA-RELATED PROTEIN.  
GN TM0727.  
OS Thermotoga maritima.  
OC Bacteria; Thermotogales; Thermotoga.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 99287316.  
RA NELSON K.E., CLAYTON R.A., GILL S.R., GWINN M.L., DODSON R.J.,  
RA HAFT D.H., HICKEY E.K., PETERSON J.D., NELSON W.C., KETCHUM K.A.,  
RA McDONALD L., UTERBACK T.R., MALEK J.A., LINHER K.D., GARRETT M.M.,  
RA STEWART A.M., COTTON M.D., PRATT M.S., PHILLIPS C.A., RICHARDSON D.,  
RA HEIDELBERG J., SUTTON G.G., FLEISCHMANN R.D., WHITE O., SALZBERG S.L.,  
RA SMITH H.O., VENTER J.C., FRASER C.M.;  
RT "Evidence for lateral gene transfer between Archaea and bacteria from  
RT genome sequence of Thermotoga maritima."  
RL Nature 399:323-329(1999).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA NELSON K.E., CLAYTON R.A., GILL S.R., GWINN M.L., DODSON R.J.,  
RA HAFT D.H., HICKEY E.K., PETERSON J.D., NELSON W.C., KETCHUM K.A.,  
RA McDONALD L., UTERBACK T.R., MALEK J.A., LINHER K.D., GARRETT M.M.,  
RA STEWART A.M., COTTON M.D., PRATT M.S., PHILLIPS C.A., RICHARDSON D.,  
RA HEIDELBERG J., SUTTON G.G., FLEISCHMANN R.D., WHITE O., SALZBERG S.L.,  
RA SMITH H.O., VENTER J.C., FRASER C.M.;  
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AE001743; AAD35809.1; -.  
SQ SEQUENCE 435 AA; 48306 MW; AD5A895D CRC32;

Query Match 71.0%; Score 49; DB 2; Length 435;  
Best Local Similarity 75.0%; Pred. No. 6.38e+00;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 350 MDRGVVIT 357  
|:|:|:|  
Qy 1 MPRGVVVT 8

RESULT 10  
ID O01487 PRELIMINARY; PRT; 470 AA.  
AC O01487;  
DT 01-JUL-1997 (TReMBLrel. 04, Created)  
DT 01-JUL-1997 (TReMBLrel. 04, Last sequence update)  
DT 01-MAY-1999 (TReMBLrel. 10, Last annotation update)  
DE COSMID C13F10.  
GN C13F10.6.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;

OC Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RX MEDLINE; 94150718;  
 RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M., COULSON A.,  
 RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., FAVELLO A.,  
 RA CRAYTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,  
 RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,  
 RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,  
 RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,  
 RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,  
 RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,  
 RA THIRRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,  
 RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;  
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
 elegans";  
 RL Nature 368:32-38(1994).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RA TIN A., WOHLDMANN P.;  
 RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; U97006; AAC47966.1; -;  
 SQ SEQUENCE 470 AA; 54820 MW; 3D5D3334F CRC32;

Query Match 71.0%; Score 49; DB 5; Length 470;  
 Best Local Similarity 44.4%; Pred. No. 6.38e+00;  
 Matches 4; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 228 MPRNLIIAL 236  
 ||| :|||  
 QY 1 MPRGVVVT 9

RESULT 11  
 ID O69875; PRELIMINARY; PRT; 648 AA.  
 AC O69875;  
 DT 01-AUG-1998 (TREMBLrel. 07, Created)  
 DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)  
 DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)  
 DE CELL DIVISION PROTEIN FTSH HOMOLOG.  
 GN FTSH.  
 OS Streptomyces coelicolor.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-A3(2);  
 RA MURPHY L., HARRIS D.;  
 RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-A3(2);  
 RA PARKHILL J., BARRELL B.G., RAJANDREAM M.A.;  
 RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-A3(2);  
 RX MEDLINE; 97000351.  
 RA REDENBACH M., KIESER H.M., DENAPATE D., EICHNER A., CULLUM J.,  
 RA KINASHI H., HOPWOOD D.A.;  
 RT "A set of ordered cosmids and a detailed genetic and physical map for  
 the 8 Mb Streptomyces coelicolor A3(2) chromosome";  
 RL Mol. Microbiol. 21:77-96(1996).  
 DR EMBL; AL023797; CAA19379.1; -;  
 DR PROSITE; PS00674; AAA; 1.  
 DR PFAM; PF00004; AAA; 1.  
 DR PFAM; PF01434; Peptidase\_M41; 1.  
 KW Cell division.  
 SQ SEQUENCE 648 AA; 69904 MW; 68219F13 CRC32;

Query Match 71.0%; Score 49; DB 2; Length 648;

Best Local Similarity 75.0%; Pred. No. 6.38e+00;  
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
 Db 235 MPRGVLLT 242  
 |||||:|  
 QY 1 MPRGVVVT 8  
 RESULT 12  
 ID O87026; PRELIMINARY; PRT; 697 AA.  
 AC O87026;  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)  
 DE CAPSID.  
 GN CAP.  
 OS Saccharomyces cerevisiae virus La.  
 OC Viruses; dsRNA viruses; Totiviridae; Totivirus.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 94111988.  
 RA BRUENN J.A.;  
 RT "A closely related group of RNA-dependent RNA polymerases from double-  
 stranded RNA viruses";  
 RL Nucleic Acids Res. 21:5667-5669(1993).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 96182949.  
 RA PARK C.M., LOPINSKI J.D., MASUDA J., TZENG T.H., BRUENN J.A.;  
 RT "A second double-stranded RNA virus from yeast";  
 RL Virology 216:451-454(1996).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA BRUENN J.A.;  
 RL Submitted (AUG-1993) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; U01060; AAB02145.1; -;  
 SQ SEQUENCE 697 AA; 78315 MW; 5C1BD5E2 CRC32;

Query Match 71.0%; Score 49; DB 14; Length 697;  
 Best Local Similarity 55.6%; Pred. No. 6.38e+00;  
 Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 393 MNRGIIVDL 401  
 ||| :|||  
 QY 1 MPRGVVVT 9

RESULT 13  
 ID O88111; PRELIMINARY; PRT; 831 AA.  
 AC O88111;  
 DT 01-NOV-1998 (TREMBLrel. 08, Created)  
 DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)  
 DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)  
 DE PERIPLASMIC NITRATE REDUCTASE PRECURSOR (EC 1.7.99.4).  
 GN NAPA.  
 OS Rhodobacter sphaeroides f. sp. denitrificans.  
 OC Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;  
 OC Rhodobacter.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-IL 106;  
 RA SAGATY M., SCHWINTNER C., CAHORS C., RICHARD P., VERMEGLIO A.;  
 RT "The periplasmic nitrate reductase of Rhodobacter sphaeroides f. sp.  
 denitrificans is essential for denitrification";  
 RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-IL106;  
 RA YAMAMOTO I.;  
 RT "Cloning of the napKEDABC genes encoding a periplasmic nitrate  
 reductase from a denitrifying phototrophic bacterium Rhodobacter  
 sphaeroides f. sp. denitrificans";  
 RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF069545; AAC23522.1; -  
DR EMBL; AB016290; BAA31961.1; -  
DR PFAM; PF00384; molybdopterin; 1.  
DR PFAM; PF01568; Molydop\_binding; 1.  
KW Signal; Oxidoreductase.  
FT SIGNAL 1  
FT CHAIN 12  
FT CHAIN 13 831  
SQ SEQUENCE 831 AA; 92711 MW; 359F61F9 CRC32;

Query Match 71.0%; Score 49; DB 2; Length 831;  
Best Local Similarity 100.0%; Pred. No. 6.38e+00;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 788 MPRGVV 793  
Qy 1 MPRGVV 6

RESULT 14  
ID Q01375 PRELIMINARY; PRT; 1154 AA.  
AC Q01375;  
DT 01-NOV-1996 (TRENBLrel. 01, Created)  
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)  
DT 01-NOV-1998 (TRENBLrel. 08, Last annotation update)  
DE HYPOTHETICAL 130.4 KD PROTEIN.  
OS Neurospora crassa.  
OC Eukaryota; Fungi; Ascomycota; Euscomycetes; Pyrenomycetes;  
OC Sordariales; Sordariaceae; Neurospora.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-J1518;  
RX MEDLINE; 94203179.  
RA CAMBARERI E.B., HELBER J., KINSEY J.A.;  
RT "Tad1-1, an active LINE-like element of Neurospora crassa.";  
RL Mol. Gen. Genet. 242:658-665(1994).  
DR EMBL; L25662; AAA21781.1; -  
DR PFAM; PF00078; rvt; 1.  
KW Hypothetical protein.  
FT DOMAIN 1019 1022 POLY-LYS.  
FT DOMAIN 1029 1034 POLY-GLU.  
SQ SEQUENCE 1154 AA; 130398 MW; DF0BA680 CRC32;

Query Match 71.0%; Score 49; DB 3; Length 1154;  
Best Local Similarity 55.6%; Pred. No. 6.38e+00;  
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 828 KPRGIVIGL 836  
Qy 1 MPRGVVVT 9

RESULT 15  
ID Q01379 PRELIMINARY; PRT; 1154 AA.  
AC Q01379;  
DT 01-NOV-1996 (TRENBLrel. 01, Created)  
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)  
DT 01-NOV-1998 (TRENBLrel. 08, Last annotation update)  
DE HYPOTHETICAL 130.5 KD PROTEIN.  
OS Neurospora crassa.  
OC Eukaryota; Fungi; Ascomycota; Euscomycetes; Pyrenomycetes;  
OC Sordariales; Sordariaceae; Neurospora.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-J1518;  
RX MEDLINE; 94203179.  
RA CAMBARERI E.B., HELBER J., KINSEY J.A.;  
RT "Tad1-1, an active LINE-like element of Neurospora crassa.";  
RL Mol. Gen. Genet. 242:658-665(1994).  
DR EMBL; L25663; AAA21792.1; -  
DR PFAM; PF00078; rvt; 1.  
KW Hypothetical protein.  
FT DOMAIN 1019 1022 POLY-LYS.  
FT DOMAIN 1029 1034 POLY-GLU.

SQ SEQUENCE 1154 AA; 130470 MW; 7FBE8EAF CRC32;  
Query Match 71.0%; Score 49; DB 3; Length 1154;  
Best Local Similarity 55.6%; Pred. No. 6.38e+00;  
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;  
Db 828 KPRGIVIGL 836  
Qy 1 MPRGVVVT 9

Search completed: Sat Apr 15 00:10:43 2000  
Job time : 105 secs.

**THIS PAGE BLANK (USPTO)**

11/12/2019 11:12:00 AM



\*\*\*\*\*

W P S R E H (TM)

\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd.

MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Sat Apr 15 00:07:53 2000; MasPar time 5.32 Seconds  
Tabular output not generated. 50.534 Million cell updates/sec

Title: >US-08-452-843-12  
Description: (1-9) from US08452843.pep  
Perfect Score: 69  
Sequence: 1 MPRGVVVT 9

Scoring table: PAM 150  
Gap 15

Searched: 82229 seqs, 29864866 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: swiss-prot38  
1:swissprot

Statistics: Mean 24.337; Variance 25.931; scale 0.939

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	56	81.2	597	1	MCM3_ENTHI DNA REPLICATION LICENS	1.14e-01
2	55	79.7	510	1	CP46_RABIT CYTOCHROME P450 4A6 PR	1.99e-01
3	53	76.8	119	1	B2MG_ATRPA BETA-2-MICROGLOBULIN P	5.95e-01
4	53	76.8	119	1	B2MG_BRAAR BETA-2-MICROGLOBULIN P	5.95e-01
5	53	76.8	119	1	B2MG_LAGLA BETA-2-MICROGLOBULIN P	5.95e-01
6	52	75.4	181	1	RK5_SYAPA CYANELLE 50S RIBOSOMAL	1.02e+00
7	51	73.9	180	1	RL5_SYN3 50S RIBOSOMAL PROTEIN	1.73e+00
8	51	73.9	511	1	CP47_RABIT CYTOCHROME P450 4A7 PR	1.73e+00
9	50	72.5	180	1	RK5_CHLVU CHLOROPLAST 50S RIBOSO	2.91e+00
10	50	72.5	445	1	GABT_MYCLE 4-AMINOBUTYRATE AMINOT	2.91e+00
11	49	71.0	65	1	TRBK_RHLSN PROBABLE CONJUGAL TRAN	4.86e+00
12	49	71.0	106	1	COLA_HORSE COLIPASE A PRECURSOR	4.86e+00
13	49	71.0	107	1	COL_RABIT COLIPASE PRECURSOR	4.86e+00
14	49	71.0	108	1	COLB_HORSE PROCOLIPASE B PRECURSO	4.86e+00
15	49	71.0	362	1	NTCP_RAT SODIUM/BILE ACID COTRA	4.86e+00
16	49	71.0	572	1	MOES_LYTVA MOESIN	4.86e+00
17	49	71.0	831	1	NAPA_PARDT PERIPLASMIC NITRATE RE	4.86e+00
18	49	71.0	831	1	NAPA_RHOSH PERIPLASMIC NITRATE RE	4.86e+00
19	48	69.6	54	1	YH1L_STRCO HYPOTHETICAL 5.6 KD PR	8.04e+00
20	48	69.6	428	1	KINC_BAGSU SPORULATION KINASE C (	8.04e+00
21	48	69.6	552	1	PUR6_SCHPO PHOSPHORIBOSYLAMINOIMI	8.04e+00
22	48	69.6	557	1	PUR6_VIGAC PHOSPHORIBOSYLAMINOIMI	8.04e+00
23	48	69.6	571	1	PUR6_YEAST PHOSPHORIBOSYLAMINOIMI	8.04e+00

24	47	68.1	95	1	COL2_PTIG COLIPASE II PRECURSOR	1.32e+01
25	47	68.1	112	1	COL_HUMAN COLIPASE PRECURSOR	1.32e+01
26	47	68.1	151	1	VG13_HSVSA IMMEDIATE EARLY GENE 1	1.32e+01
27	47	68.1	326	1	PEXA_HUMAN PEROXISOME ASSEMBLY PR	1.32e+01
28	47	68.1	334	1	TRAB_YEREN TRANSPOSOME FOR INSERT	1.32e+01
29	47	68.1	491	1	ATPB_CHLRE ATP SYNTHASE BETA CHAI	1.32e+01
30	47	68.1	1829	1	MYSD_CHICK DILUTE MYOSIN HEAVY CH	1.32e+01
31	47	68.1	1853	1	MYSA_MOUSE DILUTE MYOSIN HEAVY CH	1.32e+01
32	47	68.1	2193	1	RRPL_MEASE RNA POLYMERASE BETA SU	1.32e+01
33	47	68.1	2193	1	RRPL_MEASA RNA POLYMERASE BETA SU	1.32e+01
34	47	68.1	4543	1	LRPL_CHICK LOW-DENSITY LIPOPROTEI	1.32e+01
35	46	66.7	119	1	B2MG_SAIBB BETA-2-MICROGLOBULIN P	2.15e+01
36	46	66.7	119	1	B2MG_CACME BETA-2-MICROGLOBULIN P	2.15e+01
37	46	66.7	119	1	B2MG_CHISA BETA-2-MICROGLOBULIN P	2.15e+01
38	46	66.7	119	1	B2MG_AOTAZ BETA-2-MICROGLOBULIN P	2.15e+01
39	46	66.7	119	1	B2MG_CALGO BETA-2-MICROGLOBULIN P	2.15e+01
40	46	66.7	123	1	GALA_BOVIN GALANIN PRECURSOR	2.15e+01
41	46	66.7	187	1	RL5_MYCLE 50S RIBOSOMAL PROTEIN	2.15e+01
42	46	66.7	340	1	ARGC_STRCL N-ACETYL-GAMMA-GLUTAMY	2.15e+01
43	46	66.7	342	1	ARGC_STRCO N-ACETYL-GAMMA-GLUTAMY	2.15e+01
44	46	66.7	380	1	GALI_SALTY GALACTOKINASE (EC 2.7.	2.15e+01
45	46	66.7	543	1	PUR6_PICME PHOSPHORIBOSYLAMINOIMI	2.15e+01

ALIGNMENTS

RESULT 1	ID	MCM3_ENTHI	STANDARD;	PRT;	597 AA.
AC	024849;				
DT	01-NOV-1997	(Rel. 35, Created)			
DT	01-NOV-1997	(Rel. 35, Last sequence update)			
DE	01-NOV-1997	(Rel. 35, Last annotation update)			
DE	DNA REPLICATION LICENSING FACTOR MCM3.				
GN	MCM3.				
OS	Entamoeba histolytica.				
OC	Eukaryota; Entamoebidae; Entamoeba.				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=HM-1:IMSS;				
RA	GANGOPADHYAY S.S., LOHIA A.;				
RL	Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.				
CC	-!- FUNCTION: ACT AS A FACTOR THAT LICENSE THE DNA FOR ONE AND ONLY				
CC	ONE ROUND OF REPLICATION PER CELL CYCLE. REQUIRED FOR DNA				
CC	REPLICATION AND CELL PROLIFERATION (BY SIMILARITY).				
CC	-!- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).				
CC	-!- SIMILARITY: BELONGS TO THE MCM FAMILY.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@isb-sib.ch).				
CC	-----				
CC	EMBL; X98048; CAA66661.1; .				
DR	PROSITE; PS00847; MCM_1; 1.				
DR	PROSITE; PS0051; MCM_2; 1.				
DR	PFAM; PF00493; MCM; 1.				
KW	Transcription regulation; DNA-binding; Nuclear protein;				
KW	DNA replication; ATP-binding; Cell cycle.				
FT	DOMAIN 180 386				
FT	NP_BIND 229 236				
SQ	SEQUENCE 597 AA; 66412 MW; 3C73D4D9 CRC32;				

Query Match 81.2%; Score 56; DB 1; Length 597;

Best Local Similarity 66.7%; Pred. No. 1.14e-01;

Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 104 MPRSVIVIL 112

QY 1 MPRGVVVT 9

RESULT 2  
ID CP46\_RABIT STANDARD; PRT; 510 AA.  
AC P14580;  
DT 01-JAN-1990 (Rel. 13, Created)  
DT 01-JAN-1990 (Rel. 13, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE CYTOCHROME P450 4A6 PRECURSOR (EC 1.14.15.3) (CYP1A6) (LAURIC ACID  
DE OMEGA-HYDROXYLASE) (P450-KA-1).  
GN CYP4A6  
OS Oryctolagus cuniculus (Rabbit).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Lagomorpha; Leporidae; Oryctolagus.  
RN [1]  
RC TISSUE-KIDNEY;  
RX MEDLINE; 90254128.  
RA JOHNSON E.F., WALKER D.L., GRIFFIN K.J., CLARK J.E., OKITA R.T.,  
RA MEURHOFF A.S., MASTERS B.S.;  
RT "Cloning and expression of three rabbit kidney cDNAs encoding lauric  
RT acid omega-hydroxylases.";  
RL Biochemistry 29:873-879(1990).  
RN [2]  
RP SEQUENCE FROM N.A., AND SEQUENCE OF 5-24.  
RC TISSUE-KIDNEY;  
RX MEDLINE; 90094341.  
RA YOKOTANI M., BERNHARDT R., SOGAWA K., KUSUNOSE E., GOTOH O.,  
RA KUSUNOSE M., FUJII-KURIYAMA Y.;  
RT "Two forms of omega-hydroxylase toward prostaglandin A and laurate.  
RT cDNA cloning and their expression.";  
RL J. Biol. Chem. 264:21665-21669(1989).  
CC -1- FUNCTION: CYTOCHROMES P450 ARE A GROUP OF HEME-THIOLATE  
CC MONOOXYGENASES. IN LIVER MICROSOMES, THIS ENZYME IS INVOLVED IN AN  
CC NADPH-DEPENDENT ELECTRON TRANSPORT PATHWAY. IT OXIDIZES A VARIETY  
CC OF STRUCTURALLY UNRELATED COMPOUNDS, INCLUDING STEROIDS, FATTY  
CC ACIDS, AND XENOBIOTICS.  
CC -1- FUNCTION: THE KIDNEY P-450 SYSTEM IS RATHER SPECIALIZED FOR THE  
CC OMEGA-HYDROXYLATION OF FATTY ACIDS. P450-KA1 AND P450-KA2 CATALYZE  
CC THE OMEGA- AND (OMEGA-1)-HYDROXYLATION OF VARIOUS FATTY ACIDS WITH  
CC NO DRUG-METABOLIZING ACTIVITY, AND HYDROXYLATE PROSTAGLANDIN A1  
CC AND A2 SOLELY AT THE OMEGA-POSITION.  
CC -1- CATALYTIC ACTIVITY: OCTANE + REDUCED RUBREDOXIN + O(2) -> 1-OCTANOL  
CC + OXIDIZED RUBREDOXIN + H(2)O.  
CC -1- SUBCELLULAR LOCATION: MEMBRANE-BOUND. ENDOPLASMIC RETICULUM.  
CC -1- TISSUE SPECIFICITY: LIVER; KIDNEY.  
CC -1- INDUCTION: P450 CAN BE INDUCED TO HIGH LEVELS IN LIVER AND OTHER  
CC TISSUES BY VARIOUS FOREIGN COMPOUNDS, INCLUDING DRUGS, PESTICIDES,  
CC AND CARCINOGENS.  
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL; M28656; AAA31230.1; -  
DR EMBL; M29331; AAA31234.1; -  
DR PIR; A34160; A34160.  
DR PIR; B34260; B34260.  
DR PROSITE; PS00086; CYTOCHROME\_P450; 1.  
DR PFAM; PF00067; p450; 1.  
KW Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme;  
KW Microsome; Endoplasmic reticulum.  
FT PROPEP 1 4  
FT CHAIN 5 510 CYTOCHROME P450 4A6.  
FT BINDING 457 457 HEME.  
FT CONFLICT 424 425 VW -> CG (IN REF. 2).  
FT CONFLICT 434 434 F -> S (IN REF. 2).  
FT CONFLICT 476 476 V -> L (IN REF. 2).  
SQ SEQUENCE 510 AA; 58300 MW; E11495ED CRC32;

Query Match 79.7%; Score 53; DB 1; Length 510;  
Best Local Similarity 66.7%; Pred. No. 1.99e-01;  
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;  
Db 405 LPKGVIIVTL 413  
Oy 1 MPRGVVVT 9  
:|:|:|:|  
:|:|:|:|  
RESULT 3  
ID B2MG\_ATEPA STANDARD; PRT; 119 AA.  
AC O77536;  
DT 15-DEC-1998 (Rel. 37, Created)  
DT 15-DEC-1998 (Rel. 37, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE BETA-2-MICROGLOBULIN PRECURSOR.  
GN B2M.  
OS Ateles paniscus (Black spider monkey).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Primates; Platyrrhini; Cebidae; Ateleinae; Ateles.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 98298008.  
RA CANAVEZ F.C., LADASKY J.J., MUNIZ J.A.P.C., SEUANEZ H.N., PARHAM P.;  
RT "Beta-2-microglobulin in neotropical primates (Platyrrhini).";  
RL Immunogenetics 48:133-140(1998).  
CC -1- FUNCTION: BETA-2-MICROGLOBULIN IS THE BETA-CHAIN OF MAJOR  
CC HISTOCOMPATIBILITY COMPLEX CLASS I MOLECULES.  
CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL; AF032087; AAC52101.1; -  
DR EMBL; AF032086; AAC52101.1; JOINED.  
DR PROSITE; PS00290; IG\_MHC; 1.  
DR PFAM; PF00047; Ig; 1.  
KW MHC I; Signal.  
FT SIGNAL 1 20 BY SIMILARITY.  
FT CHAIN 21 119 BETA-2-MICROGLOBULIN.  
FT DISULFID 45 100 BY SIMILARITY.  
SQ SEQUENCE 119 AA; 13654 MW; B9EA602A CRC32;  
Query Match 76.8%; Score 53; DB 1; Length 119;  
Best Local Similarity 66.7%; Pred. No. 5.95e-01;  
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;  
Db 1 MARSVVVAL 9  
Oy 1 MPRGVVVT 9  
:|:|:|:|  
:|:|:|:|  
RESULT 4  
ID B2MG\_PRAAR STANDARD; PRT; 119 AA.  
AC O77524;  
DT 15-DEC-1998 (Rel. 37, Created)  
DT 15-DEC-1998 (Rel. 37, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE BETA-2-MICROGLOBULIN PRECURSOR.  
GN B2M.  
OS Brachyteles arachnoides (Woolly spider monkey) (Muriel).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Primates; Platyrrhini; Cebidae; Ateleinae; Brachyteles.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 98298008.  
RA CANAVEZ F.C., LADASKY J.J., MUNIZ J.A.P.C., SEUANEZ H.N., PARHAM P.;



```

RESULT 7
ID   RL5_SVNY3          STANDARD;          PRT;   180 AA.
AC   P73308;
DT   01-NOV-1997 (Rel. 35, Created)
DT   13-JUL-1998 (Rel. 36, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   50S RIBOSOMAL PROTEIN L5.
GN   RPL5 OR SLL1808.
OS   Synecocystis sp. (strain PCC 6803).
OC   Bacteria; Cyanobacteria; Chroococcales; Synecocystis.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   MEDLINE; 97061201.
RX   KANEKO T., SATO S., KOTANI H., TANAKA A., ASAMIZU E., NAKAMURA Y.,
RA   MIYAJIMA N., HIROSAWA M., SUGIURA M., SASAMOTO S., KIMURA T.,
RA   HOSOUCHI T., MATSUO A., MURAKI A., NAKAZAKI N., NARAO K.,
RA   OKUMURA S., SHIMPO S., TAKEUCHI C., WADA T., WATANABE A.,
RA   YAMADA M., YASUDA M., TABATA S.;
RT   "Sequence analysis of the genome of the unicellular cyanobacterium
RT   Synecocystis sp. strain PCC6803. II. Sequence determination of the
RT   entire genome and assignment of potential protein-coding regions.";
RL   DNA Res. 3:109-136(1996).
CC   -1- FUNCTION: THIS IS ONE OF 3 PROTEINS THAT MEDIATE THE ATTACHMENT OF
CC   THE 5S RNA INTO THE LARGE RIBOSOMAL SUBUNIT (BY SIMILARITY).
CC   -1- SIMILARITY: BELONGS TO THE L5P FAMILY OF RIBOSOMAL PROTEINS.
CC   -----
CC   THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----
DR   EMBL; D90905; BAA17337.1; ALP_INIT.
DR   PROSITE; PS00358; RIBOSOMAL_L5; 1.
DR   PFAM; PF00281; Ribosomal_L5; 1.
DR   PFAM; PF00673; Ribosomal_L5_C; 1.
KW   Ribosomal protein; rRNA-binding.
SQ   SEQUENCE 180 AA; 20230 MW; 5C38AF02 CRC32;

Query Match 73.9%; Score 51; DB 1; Length 180;
Best Local Similarity 77.8%; Pred. No. 1.73e+00;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 84 MPVGVVWTL 92
QY 1 MPRGVVWTL 9
|||:|:|
|||:|:|

RESULT 8
ID   CP47_RABIT          STANDARD;          PRT;   511 AA.
AC   P14581;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-JAN-1990 (Rel. 13, Last sequence update)
DT   15-JUL-1998 (Rel. 38, Last annotation update)
DE   CYTOCHROME P450 4A7 PRECURSOR (EC 1.14.15.3) (CYP1A7) (LAURIC ACID
DE   OMEGA-HYDROXYLASE) (P450-RA-2).
GN   CYP4A7.
OS   Oryctolagus cuniculus (Rabbit).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Lagomorpha; Leporidae; Oryctolagus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   MEDLINE; 90254128.
RX   JOHNSON E.F., WALKER D.L., GRIFFIN K.J., CLARK J.E., OKITA R.T.,
RA   MEURHOFF A.S., MASTERS B.S.;
RT   "Cloning and expression of three rabbit kidney cDNAs encoding lauric
RT   acid omega-hydroxylases".
RL   Biochemistry 29:873-879(1990).
RN   [2]

SEQUENCE FROM N.A., AND SEQUENCE OF 5-24.
TISSUE-KIDNEY;
XKOTANI N., BERNHARDT R., SOGAWA K., KUSUNOSE E., GOTOH O.,
KUSUNOSE M., FUJII-KURIYAMA Y.;
"Two forms of omega-hydroxylase toward prostaglandin A and laurate.
CDNA cloning and their expression.";
J. Biol. Chem. 264:21665-21669(1989).
-1- FUNCTION: CYTOCHROMES P450 ARE A GROUP OF HEME-THIOLATE
MONOOXYGENASES. IN LIVER MICROSOMES, THIS ENZYME IS INVOLVED IN AN
NADPH-DEPENDENT ELECTRON TRANSPORT PATHWAY. IT OXIDIZES A VARIETY
OF STRUCTURALLY UNRELATED COMPOUNDS, INCLUDING STEROIDS, FATTY
ACIDS, AND XENOBIOTICS.
-1- FUNCTION: THE KIDNEY P-450 SYSTEM IS RATHER SPECIALIZED FOR THE
OMEGA-HYDROXYLATION OF FATTY ACIDS. P450-KAL AND P450-KA2 CATALYZE
THE OMEGA- AND (OMEGA-1)-HYDROXYLATION OF VARIOUS FATTY ACIDS WITH
NO DRUG-METABOLIZING ACTIVITY, AND HYDROXYLATE PROSTAGLANDIN A1
AND A2 SOLELY AT THE OMEGA-POSITION.
-1- CATALYTIC ACTIVITY: OCTANE + REDUCED RUBREDOXIN + O(2) -> 1-OCTANOL
+ OXIDIZED RUBREDOXIN + H(2)O.
-1- SUBCELLULAR LOCATION: MEMBRANE-BOUND. ENDOPLASMIC RETICULUM.
-1- TISSUE SPECIFICITY: LIVER, KIDNEY, SMALL INTESTINE.
-1- INDUCTION: P450 CAN BE INDUCED TO HIGH LEVELS IN LIVER AND OTHER
TISSUES BY VARIOUS FOREIGN COMPOUNDS, INCLUDING DRUGS, PESTICIDES,
AND CARCINOGENS.
-1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC   -----
CC   THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----
DR   EMBL; M28657; AAA31231.1; -.
DR   EMBL; M29530; AAA31233.1; -.
DR   PIR; C34260; C34260.
DR   PIR; B34160; B34160.
DR   PROSITE; PS00086; CYTOCHROME_P450; 1.
DR   PFAM; PF00067; P450; 1.
KW   Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme;
KW   Microsome; Endoplasmic reticulum.
FT   PROPEP 1 4
FT   CHAIN 5 511 CYTOCHROME P450 4A7.
FT   BINDING 458 458 HEME.
FT   CONFLICT 99 99 C -> V (IN REF. 2).
FT   CONFLICT 150 150 F -> S (IN REF. 2).
FT   CONFLICT 392 393 SK -> RQ (IN REF. 2).
FT   CONFLICT 477 477 V -> L (IN REF. 2).
SQ   SEQUENCE 511 AA; 58318 MW; 2EA3C45E CRC32;

Query Match 73.9%; Score 51; DB 1; Length 511;
Best Local Similarity 44.4%; Pred. No. 1.73e+00;
Matches 4; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

Db 406 LPKGIIITL 414
QY 1 MPRGVVWTL 9
|||:|:|
|||:|:|

RESULT 9
ID   RK5_CHLVU          STANDARD;          PRT;   180 AA.
AC   P56362;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   13-JUL-1998 (Rel. 36, Last annotation update)
DE   CHLOROPLAST 50S RIBOSOMAL PROTEIN L5.
GN   RPL5.
OS   Chlorella vulgaris.
OC   Chloroplast.
OC   Eukaryota; Viridiplantae; Chlorophyta; Trebouxiophyceae; Chlorellales;
OC   Chlorellaceae; Chlorella.
```

RP SEQUENCE FROM N.A.  
RC STRAIN-1AM C-27 / TAMIYA;  
RX MEDLINE: 97303241.  
RA TSUDZUKI T., NAGAI T., KAPOOR M., SUGITA M., ITO M., ITO S.,  
RA WAKASUGI T., NAKASHIMA K., TSUDZUKI T., SUZUKI Y., HAMADA A., OHTA T.,  
RA INAMURA A., YOSHINAGA K., SUGIURA M.,  
RA "Complete nucleotide sequence of the chloroplast genome from the  
RT green alga *Chlorella vulgaris*: the existence of genes possibly  
RT involved in chloroplast division."  
RL Proc. Natl. Acad. Sci. U.S.A. 94:5967-5972(1997).  
CC -1- FUNCTION: THIS IS ONE OF 3 PROTEINS THAT MEDIATE THE ATTACHMENT OF  
CC THE 5S RNA INTO THE LARGE RIBOSOMAL SUBUNIT (BY SIMILARITY).  
CC -1- SIMILARITY: BELONGS TO THE LSP FAMILY OF RIBOSOMAL PROTEINS.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
DR EMBL: AB001684; BAA20745.1; -  
DR PROSITE: PS00358; RIBOSOMAL\_L5; FALSE\_NEG.  
DR PFAM: PF00281; Ribosomal\_L5; 1.  
DR PFAM: PF00673; Ribosomal\_L5\_C; 1.  
KW Ribosomal protein; rRNA-binding; Chloroplast.  
SQ SEQUENCE 180 AA; 20622 MW; E299F1E0 CRC32;  
  
Query Match 72.5%; Score 50; DB 1; Length 180;  
Best Local Similarity 77.8%; Pred. No. 2.91e+00;  
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
  
Db 84 MPVGLVWTL 92  
QY 1 MPRGVVWTL 9  
II I:IIII  
-----  
RESULT 10  
ID GABT\_MVCLC STANDARD; PRT; 446 AA.  
AC P40829;  
DT 01-FEB-1995 (Rel. 31, Created)  
DT 01-FEB-1995 (Rel. 31, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE 4-AMINOBTYRATE AMINOTRANSFERASE (EC 2.6.1.19) (GAMMA-AMINO-N-BUTYRATE  
DE TRANSAMINASE) (GABA TRANSAMINASE) (GLUTAMATE:SUCCINIC SEMIALDEHYDE  
DE TRANSAMINASE) (GABA AMINOTRANSFERASE).  
GN GABT OR B1177\_F2\_67 OR MLCB1259.03C.  
OS Mycobacterium leprae.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA SMITH D.R., ROBISON K.;  
RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.  
[2]  
CC SEQUENCE FROM N.A.  
CC DEVLIN K., CHURCHER C.M., BARRELL B.G., RAJANDREAM M.A.;  
CC Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases  
CC -1- CATALYTIC ACTIVITY: 4-AMINOBUTANOATE + 2-OXOGLUTARATE -> SUCCINATE  
CC -1- SEMIALDEHYDE + L-GLUTAMATE.  
CC -1- COFACTOR: PYRIDOXAL PHOSPHATE.  
CC -1- PATHWAY: 4-AMINOBTYRATE (GABA) DEGRADATION PATHWAY.  
CC -1- SIMILARITY: BELONGS TO CLASS-III OF PYRIDOXAL-PHOSPHATE-DEPENDENT  
CC AMINOTRANSFERASES.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
DR EMBL: U00011; AAA17107.1; -  
DR EMBL: AL023591; CAA19078.1; -  
DR HSP: P16932; IDGE.  
DR PROSITE: PS00600; AA\_TRANSFER\_CLASS\_3; 1.  
DR PFAM: PF00202; aminotran\_3; 1.  
KW Transferase; Aminotransferase; Pyridoxal phosphate.  
FT BINDING 291 291 PYRIDOXAL PHOSPHATE (BY SIMILARITY).  
SQ SEQUENCE 446 AA; 47215 MW; 7F34F294 CRC32;  
  
Query Match 72.5%; Score 50; DB 1; Length 446;  
Best Local Similarity 77.8%; Pred. No. 2.91e+00;  
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
  
Db 30 VPRGVGVTL 38  
QY 1 MPRGVVWTL 9  
IIII IIII  
-----  
RESULT 11  
ID TRBK\_RHSN STANDARD; PRT; 65 AA.  
AC P55401;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE PROBABLE CONJUGAL TRANSFER PROTEIN TRBK PRECURSOR.  
GN TRBK OR Y4DB.  
OS Rhizobium sp. (strain NGR234).  
OC Plasmid sym pNGR234a.  
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
OC Rhizobiaceae; Rhizobium.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 97305956.  
RA FREIBERG C.A., FELLAY R., BAIRICH A., BROUGHTON W.J., ROSENTHAL A.,  
RA PERRET X.;  
RA "Molecular basis of symbiosis between Rhizobium and legumes."  
RL Nature 387:394-401(1997).  
CC -1- SUBCELLULAR LOCATION: PERIPLASMIC (POTENTIAL).  
CC -1- SIMILARITY: STRONG, TO A.TUMEFACIENS TI PLASMID TRBK.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
DR EMBL: AE000068; AAB92434.1; -  
DR Conjugation; Periplasmic; Plasmid; Signal.  
FT SIGNAL 1 19 POTENTIAL.  
FT CHAIN 20 65 PROBABLE CONJUGAL TRANSFER PROTEIN TRBK.  
SQ SEQUENCE 65 AA; 6864 MW; 909C5F10 CRC32;  
  
Query Match 71.0%; Score 49; DB 1; Length 65;  
Best Local Similarity 44.4%; Pred. No. 4.86e+00;  
Matches 4; Conservative 5; Mismatches 0; Indels 0; Gaps 0;  
  
Db 1 MSRAVITL 9  
QY 1 MPRGVVWTL 9  
II:IIII  
-----  
RESULT 12  
ID COLA\_HORSE STANDARD; PRT; 106 AA.  
AC P02704;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 01-NOV-1995 (Rel. 32, Last annotation update)  
DE PROCOUPASE A PRECURSOR (FRAGMENT).  
OS Equus caballus (Horse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Perissodactyla; Equidae; Equus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-PANCREAS;  
RA MEDLINE; 94325330.  
RX CRENON I., GRANON S., CHAPUS C., KERFELEC B.;  
RT "Molecular cloning and expression of two horse pancreatic cDNA  
RL encoding colipase A and B";  
RN Biochim. Biophys. Acta 1213:357-360(1994).  
[2]  
RP SEQUENCE OF 12-106.  
RX MEDLINE; 84104937.  
RA STERNBY B., ENGSTROM A., HELLMAN U., VIHERT A.M., STERNBY N.H.,  
RA BORGSTROM B.;  
RT "The primary sequence of human pancreatic colipase";  
RL Biochim. Biophys. Acta 784:75-80(1984).  
[3]  
RP SEQUENCE OF 12-106.  
RX MEDLINE; 82186702.  
RA PIERROT M., ASTIER J.-P., ASTIER M., CHARLES M., DRENTH J.;  
RT "Pancreatic colipase: crystallographic and biochemical aspects";  
RL Eur. J. Biochem. 123:347-354(1982).  
[4]  
RP SEQUENCE OF 12-66.  
RX MEDLINE; 81021166.  
RA JULIEN R., BECHIS G., GREGOIRE J., RATHÉLOT J., ROCHAT H., SARDA L.;  
RT "Evidence for the existence of two isocolipases in horse pancreas";  
RL Biochem. Biophys. Res. Commun. 95:1243-1252(1980).  
CC -!- FUNCTION: COLIPASE IS A COFACTOR OF PANCREATIC LIPASE. IT ALLOWS  
CC THE LIPASE TO ANCHOR ITSELF TO THE LIPID-WATER INTERFACE. WITHOUT  
CC COLIPASE THE ENZYME IS WASHED OFF BY BILE SALTS, WHICH HAVE AN  
CC INHIBITORY EFFECT ON THE LIPASE.  
CC -!- FUNCTION: ENTEROSTATIN HAS A BIOLOGICAL ACTIVITY AS A SATIETY  
CC SIGNAL.  
CC -!- SUBUNIT: FORM A 1:1 STOICHIOMETRIC COMPLEX WITH PANCREATIC LIPASE.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; X74503; CAA52611.1; -.  
DR PIR; A03164; ALHQA.  
DR HSSP; P02703; IETH.  
DR PROSITE; PS00121; COLIPASE; 1.  
DR PFAM; PF01114; Colipase; 1.  
KW Lipid degradation; Digestion; Pancreas; Signal.  
FT NON\_TER 1  
FT SIGNAL <1 11  
FT PROPEP 12 16 ENTEROSTATIN, ACTIVATION PEPTIDE.  
FT CHAIN 17 106 COLIPASE A.  
FT DISULFID 28 39 BY SIMILARITY.  
FT DISULFID 34 50 BY SIMILARITY.  
FT DISULFID 38 72 BY SIMILARITY.  
FT DISULFID 60 80 BY SIMILARITY.  
FT DISULFID 74 98 BY SIMILARITY.  
FT BINDING 63 63 BILE SALT MICELLES.  
FT CONFLICT 33 33 O -> E (IN REF. 2 AND 3).  
FT CONFLICT 43 43 S -> E (IN REF. 2, 3 AND 4).  
FT CONFLICT 100 100 D -> N (IN REF. 3).  
FT CONFLICT 103 103 R -> K (IN REF. 3).  
FT CONFLICT 106 106 E -> ER (IN REF. 3).  
SQ SEQUENCE 106 AA; 11388 MW; 997D0B63 CRC32;

Query Match 71.0%; Score 49; DB 1; Length 106;  
Best Local Similarity 62.5%; Pred. No. 4.86e+00;  
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
DB 15 PRGVIINL 22

QY 2 PRGVVVT 9  
||||:|  
RESULT 13  
ID COL\_RABIT STANDARD; PRT; 107 AA.  
AC P42890;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE COLIPASE PRECURSOR.  
GN CLPS.  
OS Oryctolagus cuniculus (Rabbit).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Lagomorpha; Leporidae; Oryctolagus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-PANCREAS;  
RX MEDLINE; 93345715.  
RA COLWELL N.S., ALEMAN-GOMEZ J.A., SASSER T.L., KUMAR V.B.;  
RT "Cloning and characterization of rabbit pancreatic colipase";  
RL Int. J. Biochem. 25:885-890(1993).  
CC -!- FUNCTION: COLIPASE IS A COFACTOR OF PANCREATIC LIPASE. IT ALLOWS  
CC THE LIPASE TO ANCHOR ITSELF TO THE LIPID-WATER INTERFACE. WITHOUT  
CC COLIPASE THE ENZYME IS WASHED OFF BY BILE SALTS, WHICH HAVE AN  
CC INHIBITORY EFFECT ON THE LIPASE.  
CC -!- FUNCTION: ENTEROSTATIN HAS A BIOLOGICAL ACTIVITY AS A SATIETY  
CC SIGNAL (BY SIMILARITY).  
CC -!- SUBUNIT: FORM A 1:1 STOICHIOMETRIC COMPLEX WITH PANCREATIC LIPASE.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; L06329; AAA02911.1; -.  
DR HSSP; P02703; LPCO.  
DR PROSITE; PS00121; COLIPASE; 1.  
DR PFAM; PF01114; Colipase; 1.  
KW Lipid degradation; Digestion; Pancreas; Signal.  
FT SIGNAL 1 17  
FT PROPEP 18 22 ENTEROSTATIN, ACTIVATION PEPTIDE  
FT CHAIN 23 107 (POTENTIAL).  
FT DISULFID 34 45 COLIPASE.  
FT DISULFID 40 56 BY SIMILARITY.  
FT DISULFID 44 78 BY SIMILARITY.  
FT DISULFID 66 86 BY SIMILARITY.  
FT DISULFID 80 104 BY SIMILARITY.  
SQ SEQUENCE 107 AA; 11271 MW; 1D6F7BCE CRC32;  
Query Match 71.0%; Score 49; DB 1; Length 107;  
Best Local Similarity 62.5%; Pred. No. 4.86e+00;  
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
DB 21 PRGIVINL 28  
||||:|  
QY 2 PRGVVVT 9  
||||:|  
RESULT 14  
ID COLB\_HORSE STANDARD; PRT; 108 AA.  
AC P02705;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 01-NOV-1995 (Rel. 32, Last annotation update)  
DE PROCOLIPASE B PRECURSOR (FRAGMENT).  
OS Equus caballus (Horse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Perissodactyla; Equidae; Equus.

RP SEQUENCE FROM N.A.  
RC TISSUE-PANCREAS;  
RX MEDLINE: 94325330.  
RA CRENON I., GRANON S., CHAPUS C., KERFELEC B.;  
RT "Molecular cloning and expression of two horse pancreatic CDNA  
RL encoding colipase A and B.";  
RL Biochim. Biophys. Acta 1213:357-360(1994).  
[2]  
RP SEQUENCE OF 14-108.  
RX MEDLINE: 82046794.  
RA BONICEL J.J., COUCHOUD P.M., FOGLIZZO E., DESNUELLE P., CHAPUS C.;  
RT "Amino acid sequence of horse colipase B.";  
RL Biochim. Biophys. Acta 669:39-45(1981).  
[3]  
RP SEQUENCE OF 14-68.  
RX MEDLINE: 81021166.  
RA JULIEN R., BECHIS G., GREGOIRE J., RATHÉLOT J., ROCHAT H., SARDA L.;  
RT "Evidence for the existence of two isocollipases in horse pancreas.";  
RL Biochem. Biophys. Res. Commun. 95:1245-1252(1980).  
CC -1- FUNCTION: COLIPASE IS A COFACTOR OF PANCREATIC LIPASE. IT ALLOWS  
CC THE LIPASE TO ANCHOR ITSELF TO THE LIPID-WATER INTERFACE. WITHOUT  
CC COLIPASE THE ENZYME IS WASHED OFF BY BILE SALTS, WHICH HAVE AN  
CC INHIBITORY EFFECT ON THE LIPASE.  
CC -1- FUNCTION: ENTEROSTATIN HAS A BIOLOGICAL ACTIVITY AS A SATIETY  
CC SIGNAL.  
CC -1- SUBUNIT: FORM A 1:1 STOICHIOMETRIC COMPLEX WITH PANCREATIC LIPASE.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL; X74344; CAA52391.1; -.  
DR PIR; A03165; XLHOB.  
DR HSSP; P02703; IETH.  
DR PROSITE; PS00121; COLIPASE: 1.  
DR PFAM; PF01114; Colipase: 1.  
KW Lipid degradation; Digestion; Pancreas; Signal.  
FT NON\_TER 1  
FT SIGNAL <1 13  
FT PROPEP 14 18 ENTEROSTATIN, ACTIVATION PEPTIDE.  
FT CHAIN 19 108 COLIPASE B.  
FT BINDING 65 65 BILE SALT MICELLES.  
FT DISULFID 30 41 BY SIMILARITY.  
FT DISULFID 36 52 BY SIMILARITY.  
FT DISULFID 40 74 BY SIMILARITY.  
FT DISULFID 62 82 BY SIMILARITY.  
FT DISULFID 76 100 BY SIMILARITY.  
FT CONFLICT 35 35 Q -> E (IN REF. 3).  
FT CONFLICT 42 42 H -> T (IN REF. 3).  
FT CONFLICT 108 108 E -> ER (IN REF. 2).  
SQ SEQUENCE 108 AA; 11618 MW; 1A17861D CRC32;

Query Match 71.0%; Score 49; DB 1; Length 108;  
Best Local Similarity 62.5%; Pred. No. 4.86e+00;  
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 17 PRGVIIINL 24  
QY 2 PRGVVVTL 9

RESULT 15  
ID NTCP\_RAT STANDARD; PRT; 362 AA.  
AC P26435;  
DT 01-AUG-1992 (Rel. 23, Created)  
DT 01-AUG-1992 (Rel. 23, Last sequence update)  
DT 01-FEB-1994 (Rel. 28, Last annotation update)  
DE SODIUM/BILE ACID COTRANSPORTER (NA(+)/BILE ACID COTRANSPORTER)

DE (NA(+)/TAUROCHOLATE TRANSPORT PROTEIN) (SODIUM/TAUROCHOLATE  
DE COTRANSPORTING POLYPEPTIDE).  
GN SLC10A1 OR NTCP.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;  
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=LIVER;  
RX MEDLINE: 92073340.  
RA HAGENBUCH B., STIEGER B., FOGUET M., LUEBBERT H., MEIER P.J.;  
RT "Functional expression cloning and characterization of the hepatocyte  
RL Na+/bile acid cotransport system.";  
RL Proc. Natl. Acad. Sci. U.S.A. 88:10629-10633(1991).  
CC -1- FUNCTION: THE HEPATIC SODIUM/BILE ACID UPTAKE SYSTEM EXHIBITS  
CC BROAD SUBSTRATE SPECIFICITY & TRANSPORTS VARIOUS NONBILE ACID  
CC ORGANIC COMPOUNDS AS WELL. IT IS STRICTLY DEPENDENT ON THE  
CC EXTRACELLULAR PRESENCE OF SODIUM.  
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.  
CC -1- TISSUE SPECIFICITY: LIVER AND KIDNEY.  
CC -1- SIMILARITY: BELONGS TO THE SODIUM:BILE ACID SYMPORTER FAMILY  
CC (SBE).  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL; M77479; AAA42112.1; -.  
DR PIR; A41601; A41601.  
KW Transmembrane; Transport; Symport; Sodium transport; Glycoprotein.  
FT TRANSMEM 24 45 POTENTIAL.  
FT TRANSMEM 60 80 POTENTIAL.  
FT TRANSMEM 82 98 POTENTIAL.  
FT TRANSMEM 158 178 POTENTIAL.  
FT TRANSMEM 190 211 POTENTIAL.  
FT TRANSMEM 228 244 POTENTIAL.  
FT TRANSMEM 285 306 POTENTIAL.  
FT CARBOHYD 5 5 POTENTIAL.  
FT CARBOHYD 11 11 POTENTIAL.  
FT CARBOHYD 103 103 POTENTIAL.  
FT CARBOHYD 117 117 POTENTIAL.  
FT CARBOHYD 271 271 POTENTIAL.  
SQ SEQUENCE 362 AA; 39295 MW; 69E1D9DC CRC32;

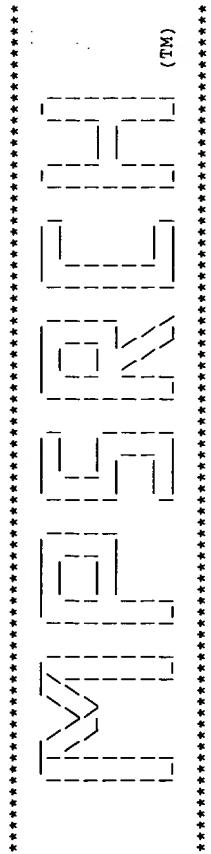
Query Match 71.0%; Score 49; DB 1; Length 362;  
Best Local Similarity 55.6%; Pred. No. 4.86e+00;  
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 57 KPKGIVVAL 65  
QY 1 MPRGVVVTL 9

Search completed: Sat Apr 15 00:08:39 2000  
Job time : 46 secs.

THIS PAGE BLANK (USPTO)





Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Sat Apr 15 00:12:27 2000; MasPar time 5.56 Seconds  
Tabular output not generated. 42.628 Million cell updates/sec

Title: >US-08-452-843-13  
Description: (1-10) from US08452843.pap  
Perfect Score: 67  
Sequence: 1 LPENNVLSP 10

Scoring table: PAM 150  
Gap 15

Searched: 188963 seqs, 23686106 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: a-geneseq36  
1:geneseq

Statistics: Mean 16.262; Variance 52.652; scale 0.309

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES				Pred. No.	
Result No.	Score	Query Match	Description	ID	
1	67	100.0	Cytotoxic T lymphocyte	11	1.88e+00
2	67	100.0	Immunodominant epitope	15	1.88e+00
3	67	100.0	Immunodominant epitope	25	1.88e+00
4	67	100.0	Human MDM2 binding p53	64	1.88e+00
5	67	100.0	Human p53 N-terminal f	64	1.88e+00
6	67	100.0	Human p53, involved in	64	1.88e+00
7	67	100.0	Human p53 amino acids	241	1.88e+00
8	67	100.0	Chimeric p53 protein.	337	1.88e+00
9	67	100.0	Del356-393 modified hu	355	1.88e+00
10	67	100.0	Chimeric p53 protein.	359	1.88e+00
11	67	100.0	Modified p53 variant p	363	1.88e+00
12	67	100.0	Modified p53 variant p	363	1.88e+00
13	67	100.0	Modified p53 variant p	363	1.88e+00
14	67	100.0	Modified p53 variant p	363	1.88e+00
15	67	100.0	Modified p53 variant p	363	1.88e+00
16	67	100.0	Modified p53 variant p	363	1.88e+00
17	67	100.0	Modified p53 variant p	363	1.88e+00
18	67	100.0	Modified p53 variant p	363	1.88e+00
19	67	100.0	Amino acid sequence of	393	1.88e+00
20	67	100.0	Human p53 mutant 1.	393	1.88e+00
21	67	100.0	Human wild-type p53 pr	393	1.88e+00
22	67	100.0	Human p53 protein SEQ	393	1.88e+00
23	67	100.0	Human p53 used in coup	393	1.88e+00

24	67	100.0	393	1	W57244	Human p53 protein SEQ	1.88e+00
25	67	100.0	393	1	W05346	Human p53 mutant R273H	1.88e+00
26	67	100.0	393	1	W05347	Human p53 mutant R248Q	1.88e+00
27	67	100.0	393	1	W13968	Modified p53 variant p	1.88e+00
28	67	100.0	393	1	W13970	Modified p53 variant p	1.88e+00
29	67	100.0	393	1	W13953	T284K modified human p	1.88e+00
30	67	100.0	393	1	W13980	Human tumour-derived p	1.88e+00
31	67	100.0	393	1	W13981	Human tumour-derived p	1.88e+00
32	67	100.0	393	1	R91933	Wild type p53 protein.	1.88e+00
33	67	100.0	393	1	R94623	p53 protein.	1.88e+00
34	67	100.0	393	1	R26758	p53	1.88e+00
35	67	100.0	393	1	W13949	T284K modified human p	1.88e+00
36	67	100.0	393	1	W02617	Human p53 tumour suppr	1.88e+00
37	67	100.0	393	1	W13951	Human tumour-derived p	1.88e+00
38	67	100.0	393	1	W13948	Human wild-type p53 cu	1.88e+00
39	67	100.0	393	1	W13979	Human tumour-derived p	1.88e+00
40	67	100.0	393	1	W05349	Human p53 mutant R273C	1.88e+00
41	67	100.0	402	1	W13965	Chimeric p53 protein.	1.88e+00
42	67	100.0	404	1	W13963	Chimeric p53 protein.	1.88e+00
43	67	100.0	406	1	W13964	Chimeric p53 protein.	1.88e+00
44	67	100.0	438	1	R74272	Tumour suppressor prot	1.88e+00
45	67	100.0	533	1	W19763	p53-GM-CSF immunostimu	1.88e+00

ALIGNMENTS

RESULT 1

ID R97509 standard; peptide; 11 AA.

AC R97509;

DT 11-FEB-1997 (first entry)

DE Cytotoxic T lymphocyte-activating peptide, corresp. to p53 aa 25-35.

KW p53; Her-2; Neu; aa; amino acid; CTL; cytotoxic T lymphocyte; target;

OS Homo sapiens.

PN W09618409-A1.

PD 20-JUN-1996.

PR 14-DEC-1995; U16415.

PF 14-DEC-1994; US-355558.

PA (SCRI ) SCRIPPS RES INST.

PI Sherman LA:

DR WPI: 96-300385/30.

PT In vivo activation of tumour-specific cytotoxic T lymphocytes - by

PT contacting with polypeptide(s) derived from human p53 or Her-2/Neu

PT proteins

PS DCIaim 40; Page 73; 158pp; English.

CC R97509 is a peptide capable of activating cytotoxic T lymphocytes

CC (CTLs) which specifically target malignant cells. The peptide

CC corresponds to amino acids 25-35 of human p53 protein. CTL-

CC activating peptides can be used in a vaccine for protecting against

CC tumour cell formation. CTLs activated by the peptides will lyse

CC tumour cells displaying specific peptides. Antibodies against CTL-

CC activating peptides are useful for the identification of other

CC similar compounds which may be useful for treating cancer or virally-

CC infected cells, or for diagnosis. The peptide and vaccines produced

CC provide immunity to a high percentage of different ethnic groups,

CC i.e. those with different HLA alleles.

SC Sequence 11 AA.

Query Match 100.0%; Score 67; DB 1; Length 11;  
Best Local Similarity 100.0%; Pred. No. 1.88e+00;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db	2	LPENNVLSP 11			
Qy	1	LPENNVLSP 10			
RESULT 2					
ID	R54911	standard; peptide; 15 AA.			
AC	R54911;				
DT	29-NOV-1994 (first entry)				
DE	Immunodominant epitope from p53 N-terminal.				
KW	cancer; pre-cancerous state; detection; diagnosis; human p53 gene;				

KW Immunodominant epitope; human cellular tumour antigen;  
 KW transformation-associated protein.  
 OS Homo sapiens.  
 PN W09410306-A.  
 PD 11-MAY-1994.  
 PF 02-NOV-1993; F01082.  
 PR 02-NOV-1992; FR-013110.  
 PA (EURO-) LAB EURO BIO SA.  
 PI Legros Y, Lubin R, Soussi T;  
 DR WPI; 94-167463/20.  
 PT New immuno:dominant epitope(s) of protein p53 - for detecting and  
 PT monitoring antibodies indicative of cancer and precancerous  
 PT states  
 PS Claim 5; Page 42; 62pp; French.  
 CC Peptides derived from the N-terminal (amino acids 1-112) or the C-  
 CC terminal (amino acids 350-393) of protein p53 which specifically  
 CC react with anti-p53 antibodies in patients with cancer or  
 CC precancerous conditions are claimed. The peptides (R54907-R54921)  
 CC are useful for detecting and monitoring cancerous and precancerous  
 CC conditions.  
 SQ Sequence 15 AA;  
 Query Match 100.0%; Score 67; DB 1; Length 15;  
 Best Local Similarity 100.0%; Pred. No. 1.88e+00;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 6 LPENNVLSP 15  
 | | | | | | | |  
 QY 1 LPENNVLSP 10  
 | | | | | | | |  
 RESULT 3  
 ID R54907 standard; peptide; 25 AA.  
 AC R54907;  
 DT 29-NOV-1994 (first entry)  
 DE Immunodominant epitope from p53 N-terminal.  
 KW cancer; pre-cancerous state; detection; diagnosis; human p53 gene;  
 KW immunodominant epitope; human cellular tumour antigen;  
 KW transformation-associated protein.  
 OS Homo sapiens.  
 PN W09410306-A.  
 PD 11-MAY-1994.  
 PF 02-NOV-1993; F01082.  
 PR 02-NOV-1992; FR-013110.  
 PA (EURO-) LAB EURO BIO SA.  
 PI Legros Y, Lubin R, Soussi T;  
 DR WPI; 94-167463/20.  
 PT New immuno:dominant epitope(s) of protein p53 - for detecting and  
 PT monitoring antibodies indicative of cancer and precancerous  
 PT states  
 PS Claim 4; Page 42; 62pp; French.  
 CC Peptides derived from the N-terminal (amino acids 1-112) or the C-  
 CC terminal (amino acids 350-393) of protein p53 which specifically  
 CC react with anti-p53 antibodies in patients with cancer or  
 CC precancerous conditions are claimed. The peptides (R54907-R54921)  
 CC are useful for detecting and monitoring cancerous and precancerous  
 CC conditions.  
 SQ Sequence 25 AA;  
 Query Match 100.0%; Score 67; DB 1; Length 25;  
 Best Local Similarity 100.0%; Pred. No. 1.88e+00;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 16 LPENNVLSP 25  
 | | | | | | | |  
 QY 1 LPENNVLSP 10  
 | | | | | | | |  
 RESULT 4  
 ID W94303 standard; protein; 64 AA.  
 AC W94303;  
 DT 13-APR-1999 (first entry)  
 DE Human MDM2 binding p53 fragment.

KW Human; MDM2; p53; tumorigenesis; growth regulation; diagnosis;  
 KW malignant fibrous histiocytoma; MFH; liposarcoma.  
 OS Homo sapiens.  
 PN US5858976-A.  
 PD 12-JAN-1999.  
 PF 14-FEB-1997; 801718.  
 PR 07-APR-1993; US-044619.  
 PR 07-APR-1992; US-867840.  
 PR 23-JUN-1992; US-903103.  
 PR 17-FEB-1995; US-390515.  
 PR 14-FEB-1997; US-801718.  
 PA (UYJO ) UNIV JOHNS HOPKINS.  
 PI Burrell M, Hill DE, Kinzler KW, Vogelstein B;  
 DR WPI; 99-152105/13.  
 PT Inhibiting growth of tumour cells having MDM2 gene amplification -  
 PT with MDM2-binding p53 fragment  
 PS Claim 1: Column 19-20; 41pp; English.  
 CC The present invention describes: (1) a method for inhibiting the growth  
 CC of tumour cells which contain a human MDM2 gene amplification,  
 CC comprising administering to the cells a DNA molecule that expresses a  
 CC polypeptide consisting of a portion of p53 i.e. amino acids 13-41 of the  
 CC present 64 amino acid sequence, the polypeptide being capable of binding  
 CC to human MDM2 (see W94304); (2) a method as in (1) where the polypeptide  
 CC lacks the homo-oligomerisation domain of p53; and (3) a method as in (1)  
 CC where the polypeptide lacks amino acids 138-393 of p53. The method is  
 CC useful for treating the following tumour types which have a MDM2 gene  
 CC amplification: M-7 malignant fibrous histiocytoma (MFH), M-20 MFH, L-9  
 CC liposarcoma, KL7 liposarcoma, KL28 liposarcoma, KL30 liposarcoma, and  
 CC OSA-CL MFH.  
 SQ Sequence 64 AA;  
 Query Match 100.0%; Score 67; DB 1; Length 64;  
 Best Local Similarity 100.0%; Pred. No. 1.88e+00;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 26 LPENNVLSP 35  
 | | | | | | | |  
 QY 1 LPENNVLSP 10  
 | | | | | | | |  
 RESULT 5  
 ID W57240 standard; protein; 64 AA.  
 AC W57240;  
 DT 10-AUG-1998 (first entry)  
 DE Human p53 N-terminal fragment.  
 KW Human; p53; MDM2; tumour; growth inhibition; amplification;  
 KW malignant fibrous histiocytoma; liposarcoma.  
 OS Homo sapiens.  
 PN US5756455-A.  
 PD 26-MAY-1998.  
 PF 17-FEB-1995; 390515.  
 PR 07-APR-1993; US-044619.  
 PR 07-APR-1992; US-867840.  
 PR 23-JUN-1992; US-903103.  
 PR 17-FEB-1995; US-390515.  
 PA (UYJO ) UNIV JOHNS HOPKINS.  
 PI Kinzler KW, Vogelstein B;  
 DR WPI; 98-321574/28.  
 PT Inhibiting growth of tumour cells having MDM2 gene amplification -  
 PT with p53 protein fragment  
 PS Claim 1: Column 19-20; 40pp; English.  
 CC A method has been developed for inhibiting the growth of tumour cells  
 CC containing a human MDM2 gene amplification. The method comprises  
 CC treating the tumour cells with a DNA molecule that expresses a  
 CC polypeptide capable of binding to human MDM2 protein. The present  
 CC sequence represents an N-terminal fragment of p53 which can bind to the  
 CC human MDM2 protein. The present invention describes three preferred  
 CC polypeptides for binding human MDM2: (1) the polypeptide comprises  
 CC amino acids 1-50 of p53 (see W57240); (2) the polypeptide comprises  
 CC amino acids 13-41 of p53 (see W57240) and at least none additional p53  
 CC residues on the N- or C-terminal side, provided that the polypeptide  
 CC lacks the homo-oligomerisation domain of p53; (3) the polypeptide  
 CC comprises amino acids 13-41 of p53 (see W57240) and at least nine

CC additional p53 residues on the N- or C-terminal side, provided that the  
 CC polypeptide lacks amino acids 138-393 of p53. Some malignant fibrous  
 CC histiocytomas and liposarcomas have an MDM2 gene amplification, so  
 CC detection of increased expression of MDM2 gene products indicates  
 CC tumorigenesis.  
 SQ Sequence 64 AA;

Query Match 100.0%; Score 67; DB 1; Length 64;  
 Best Local Similarity 100.0%; Pred. No. 1.88e+00;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 26 LPENNVLSPL 35  
 | | | | | | | | | |  
 QY 1 LPENNVLSPL 10

## RESULT 6

ID W07886 standard; protein; 64 AA.

AC W07886; 1997 (first entry)  
 DE Human p53, involved in tumour suppression.  
 KW p53; MDM-2; binding-inhibitor; identification; tumour; cancer;  
 KW neoplasia; antibody fusion protein; therapy.  
 OS Homo sapiens.

FT Key Location/Qualifiers  
 FT region 1..41  
 FT /note= "MDM-2 binding fragment"  
 FT region 1..50  
 FT /note= "MDM-2 binding fragment"  
 FT region 13..57  
 FT /note= "MDM-2 binding fragment"

PN US5550023-A.

PD 27-AUG-1996.  
 PF 07-APR-1992; 867840.  
 PR 07-APR-1992; US-867840.  
 PR 23-JUN-1992; US-903103.  
 PR 07-APR-1993; US-044619.  
 PR 18-MAY-1994; US-245500.  
 PA (UJJO) UNIV JOHNS HOPKINS.  
 PI Kinzler KW, Vogelstein B;  
 WPI: 96-401591/40.

PT Identification of cpds, interfering with human MDM2/p53 binding -  
 PT useful as therapeutic agents to treat human neoplastic cells  
 PS Claim 16; Column 19-20; 36pp; English.  
 CC W07886 represents the human p53 protein which is involved in the  
 CC for identification of many cancers. The protein is used here in a method  
 CC for identifying compounds that interfere with the binding of p53 and  
 CC MDM-2. In binding the p53 protein, the MDM-2 protein releases a cell  
 CC from p53-regulated growth, allowing cancers to develop. Therefore  
 CC compounds identified as interfering with the binding of MDM-2 to p53  
 CC are potentially useful in the treatment of human neoplastic cells. In  
 CC the method pref. one or both of the proteins is a fusion protein esp.  
 CC with an antibody or antibody fragment which aids separation and  
 CC identification  
 SQ Sequence 64 AA;

Query Match 100.0%; Score 67; DB 1; Length 64;  
 Best Local Similarity 100.0%; Pred. No. 1.88e+00;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 26 LPENNVLSPL 35  
 | | | | | | | | | |  
 QY 1 LPENNVLSPL 10

## RESULT 7

ID R51872 standard; Protein; 241 AA.

AC R51872;  
 DT 18-NOV-1994 (first entry)  
 DE Human p53 amino acids 1-241.  
 KW Human nuclear phosphoprotein p53; tumour suppressor gene product;  
 KW anti-oncogene; cancer; tumour; antibody binding region; epitope.  
 OS Homo sapiens.

PN W09408241-A.  
 PD 14-APR-1994.  
 PF 30-SEP-1993; E02666.  
 PR 30-SEP-1992; DE-232823.  
 PA (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM STIFTUNG.  
 PI Klein R, Schranz P, Tesser C, Volkman M, Zentgraf H;  
 WPI: 94-135732/16.  
 DR N-PSDB; Q62357.  
 PT Non-radioactive detection of p53 specific antibodies - by capture  
 PT on immobilised p53 or its fragments, then reaction with labelled  
 PT second antibody, for diagnosis of tumours and suitable for  
 PT screening  
 PS Claim 10; Page 17; 35pp; German.  
 CC Antibodies specific for p53 are detected by binding to immobilised  
 CC fragments of the p53 gene product containing the antibody-binding  
 CC region. Preferred fragments contain amino acids 1-241, 40-349,  
 CC 40-393, 66-241, 66-393, 237-349, 237-393 and esp. 9-33, 37-52 or  
 CC 368-386. See R51872-R51881 for sequences of these fragments.  
 SQ Sequence 241 AA;

Query Match 100.0%; Score 67; DB 1; Length 241;  
 Best Local Similarity 100.0%; Pred. No. 1.88e+00;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 26 LPENNVLSPL 35  
 | | | | | | | | | |  
 QY 1 LPENNVLSPL 10

## RESULT 8

ID W13962 standard; Protein; 337 AA.

AC W13962;  
 DT 25-JUN-1997 (first entry)  
 DE Chimeric p53 protein.  
 KW p53; tumour suppressor; cancer; therapy; cell proliferation;  
 KW apoptosis; protein engineering; GCN4; DNA binding.  
 OS Chimeric Homo sapiens;  
 OS Chimeric synthetic.

FT Key Location/Qualifiers  
 FT region 1..300  
 FT /label= p53wt  
 FT /note= "amino acids 1-300 of wild-type p53"  
 FT region 301..305  
 FT /label= Linker  
 FT region 306..337  
 FT /label= GCN4  
 FT /note= "amino acids 250-281 of GCN4 LZ variant"

PN W09710843-A1.

PD 27-MAR-1997.  
 PF 20-SEP-1996; U15188.  
 PR 22-SEP-1995; US-004802.  
 PR 21-AUG-1996; US-697221.  
 PA (WIST-) WISTAR INST ANATOMY & BIOLOGY.  
 PI Halazonetis TD;

DR WPI: 97-202618/18.  
 PT R284K modified p53 protein having DNA binding ability - useful in  
 PT treatment of cancer  
 PS Disclosure: Refer to Page 8; 82pp; English.  
 CC Chimeric p53 constructs (W13956-67) comprise N-terminal portions  
 CC of human wild-type p53 tumour suppressor (see also W13948) linked  
 CC to a C-terminal portion of the LZ variant (see also W13955) of  
 CC GCN4 and, in some cases, the C-terminal portion of wild-type  
 CC p53. The chimeric proteins have DNA binding activity and can  
 CC replace lost or insufficient p53 function, providing the means for  
 CC pharmacological rescue of p53 function in cancer patients. Nucleic  
 CC acids coding for modified p53 constructs can be used for cancer  
 CC gene therapy.  
 SQ Sequence 337 AA;

Query Match 100.0%; Score 67; DB 1; Length 337;  
 Best Local Similarity 100.0%; Pred. No. 1.88e+00;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 26 LPENNVLSP 35  
 |||||  
 QY 1 LPENNVLSP 10

## RESULT 9

ID W13950 standard; Protein; 355 AA.  
 AC W13950;  
 DT 25-JUN-1997 (first entry)  
 DE Del356-393 modified human p53.  
 KW p53; tumour suppressor; cancer; therapy; cell proliferation;  
 KW apoptosis; protein engineering; DNA binding.  
 OS Synthetic.  
 PN WO9710843-A1.  
 PD 27-MAR-1997.  
 PF 20-SEP-1996; U15188.  
 PR 22-SEP-1995; US-004802.  
 PR 21-AUG-1996; US-697221.  
 PA (WIST-) WISTAR INST ANATOMY & BIOLOGY.  
 PI Halazonetis TD;  
 DR WPI; 97-202618/18.  
 PT R284K modified p53 protein having DNA binding ability - useful in  
 treatment of cancer  
 PS Claim 3; Refer to Page 27-29; 82pp; English.  
 CC Del356-393 modified p53 (W13950) has the C-terminal region of  
 CC wild-type human p53 tumour suppressor (W13948) deleted. Modified  
 CC p53 constructs (see also W13954, W13956-61, W13971-77) bearing  
 CC a deletion of all or a fragment of the C-terminal residues  
 CC 356-393 have DNA binding ability and can activate the DNA binding  
 CC of common Class I p53 tumour mutants (see also W13951-52). The  
 CC method provides the means for pharmacological rescue of p53  
 CC function in cancer patients. Nucleic acids coding for such  
 CC constructs can be used for cancer gene therapy.  
 SQ Sequence 355 AA;

Query Match 100.0%; Score 67; DB 1; Length 355;  
 Best Local Similarity 100.0%; Pred. No. 1.88e+00; Indels 0; Gaps 0;  
 Matches 10; Conservative 0; Mismatches 0;

Db 26 LPENNVLSP 35  
 |||||  
 QY 1 LPENNVLSP 10

## RESULT 10

ID W13960 standard; Protein; 359 AA.  
 AC W13960;  
 DT 25-JUN-1997 (first entry)  
 DE Chimeric p53 protein.  
 KW p53; tumour suppressor; cancer; therapy; cell proliferation;  
 KW apoptosis; protein engineering; GCN4; DNA binding.  
 OS Chimeric Homo sapiens;  
 OS Chimeric synthetic.  
 FH Key Location/Qualifiers  
 FT region 1..323  
 FT /label= p53wt  
 FT /note= "amino acids 1-323 of wild-type p53"  
 FT region 324..326  
 FT /label= Linker  
 FT region 327..359  
 FT /label= GCN4  
 FT /note= "amino acids 249-281 of GCN4 L2 variant"  
 PN WO9710843-A1.  
 PD 27-MAR-1997.  
 PF 20-SEP-1996; U15188.  
 PR 22-SEP-1995; US-004802.  
 PR 21-AUG-1996; US-697221.  
 PA (WIST-) WISTAR INST ANATOMY & BIOLOGY.  
 PI Halazonetis TD;  
 DR WPI; 97-202618/18.  
 PT R284K modified p53 protein having DNA binding ability - useful in  
 treatment of cancer  
 PS Disclosure; Refer to Page 8; 82pp; English.

CC Chimeric p53 constructs (W13956-67) comprise N-terminal portions  
 CC of human wild-type p53 tumour suppressor (see also W13948) linked  
 CC to a C-terminal portion of the L2 variant (see also W13955) of  
 CC GCN4 and, in some cases, the C-terminal portion of wild-type  
 CC p53. The chimeric proteins have DNA binding activity and can  
 CC replace lost or insufficient p53 function, providing the means for  
 CC pharmacological rescue of p53 function in cancer patients. Nucleic  
 CC acids coding for modified p53 constructs can be used for cancer  
 CC gene therapy.  
 SQ Sequence 359 AA;

Query Match 100.0%; Score 67; DB 1; Length 359;  
 Best Local Similarity 100.0%; Pred. No. 1.88e+00; Indels 0; Gaps 0;  
 Matches 10; Conservative 0; Mismatches 0;

Db 26 LPENNVLSP 35  
 |||||  
 QY 1 LPENNVLSP 10

## RESULT 11

ID W13974 standard; Protein; 363 AA.  
 AC W13974;  
 DT 25-JUN-1997 (first entry)  
 DE Modified p53 variant p53H273del364-393.  
 KW p53; tumour suppressor; cancer; therapy; cell proliferation;  
 KW apoptosis; protein engineering; DNA binding.  
 OS Synthetic.  
 PN WO9710843-A1.  
 PD 27-MAR-1997.  
 PF 20-SEP-1996; U15188.  
 PR 22-SEP-1995; US-004802.  
 PR 21-AUG-1996; US-697221.  
 PA (WIST-) WISTAR INST ANATOMY & BIOLOGY.  
 PI Halazonetis TD;  
 DR WPI; 97-202618/18.  
 PT R284K modified p53 protein having DNA binding ability - useful in  
 treatment of cancer  
 PS Example 1; 56-57; 82pp; English.  
 CC Modified p53 variant p53H273del364-393 (W13974) has the tumour-  
 CC derived histidine 273 mutation (see also W13952) and a deletion  
 CC of the C-terminal 30 amino acids of wild-type p53 (see also  
 CC W13948). His273 is a Class I p53 tumour mutation that affects DNA  
 CC binding. The C-terminal deletion, introduced by site-directed  
 CC mutagenesis of p53 DNA, activates the DNA binding of the p53  
 CC tumour mutant. This provides the means for pharmacological rescue  
 CC of p53 function in cancer patients. Other modified p53 constructs  
 CC (W13949-50, W13953-54, W13968-77) have also been produced. Nucleic  
 CC acids coding for modified p53 can be used for cancer gene therapy.  
 SQ Sequence 363 AA;

Query Match 100.0%; Score 67; DB 1; Length 363;  
 Best Local Similarity 100.0%; Pred. No. 1.88e+00;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 26 LPENNVLSP 35  
 |||||  
 QY 1 LPENNVLSP 10

## RESULT 12

ID W13973 standard; Protein; 363 AA.  
 AC W13973;  
 DT 25-JUN-1997 (first entry)  
 DE Modified p53 variant p53Q248R284del364-393.  
 KW p53; tumour suppressor; cancer; therapy; cell proliferation;  
 KW apoptosis; protein engineering; DNA binding.  
 OS Synthetic.  
 PN WO9710843-A1.  
 PD 27-MAR-1997.  
 PF 20-SEP-1996; U15188.  
 PR 22-SEP-1995; US-004802.  
 PR 21-AUG-1996; US-697221.

PA (WIST-) WISTAR INST ANATOMY & BIOLOGY.  
 PI Halazonetis TD;  
 DR WPI; 97-202618/18.  
 PT R284K modified p53 protein having DNA binding ability - useful in  
 PT treatment of cancer  
 PS Example 1: 54-56; 82pp; English.  
 CC Modified p53 variant p53C273R284del364-393 (W13973) has the tumour-  
 CC derived Gln248 mutation. (see also W13951), a Thr284 to Arg substn.  
 CC (see also W13949) and a deletion of the 30 C-terminal amino acids  
 CC of wild-type p53 (W13948). Gln248 is a Class I p53 tumour mutation  
 CC that affects DNA binding. The T284R substitution, introduced by  
 CC site-directed mutagenesis of p53 DNA, provides a novel p53-DNA  
 CC contact between a phosphate of the DNA backbone and p53, and  
 CC restores DNA binding. The C-terminal deletion permits in vitro  
 CC DNA binding. The construct provides the means for pharmacological  
 CC rescue of p53 function in cancer patients. Other modified p53  
 CC constructs (W13949-50, W13953-54, W13968-77) have also been  
 CC produced. Nucleic acids coding for modified p53 can be used for  
 CC cancer gene therapy.  
 SQ Sequence 363 AA;

Query Match 100.0%; Score 67; DB 1; Length 363;  
 Best Local Similarity 100.0%; Pred. No. 1.88e+00;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 26 LPENNVSPL 35  
 |||||  
 QY 1 LPENNVSPL 10

## RESULT 13

ID W13971 standard; Protein; 363 AA.  
 AC W13971;  
 DT 25-JUN-1997 (first entry)  
 DE Modified p53 variant p53R284del364-393.  
 KW p53; tumour suppressor; cancer; therapy; cell proliferation;  
 KW apoptosis; protein engineering; DNA binding.  
 OS Synthetic.  
 PN W09710843-Al.  
 PD 27-MAR-1997.  
 PF 20-SEP-1996; U15188.  
 PR 22-SEP-1995; US-004802.  
 PR 21-AUG-1996; US-697221.  
 PA (WIST-) WISTAR INST ANATOMY & BIOLOGY.  
 PI Halazonetis TD;  
 DR WPI; 97-202618/18.  
 PT R284K modified p53 protein having DNA binding ability - useful in  
 PT treatment of cancer  
 PS Example 1: 51-52; 82pp; English.  
 CC Modified p53 variant p53R284del364-393 (W13971) has a Thr284 to Arg  
 CC substn. (see also W13949) and a deletion of the C-terminal 30  
 CC amino acids. The T284R substitution, introduced by site-directed  
 CC mutagenesis of p53 DNA, provides a novel p53-DNA contact between a  
 CC phosphate of the DNA backbone and p53. The C-terminal deletion  
 CC permits in vitro DNA binding. The variant provides the means for  
 CC pharmacological rescue of p53 function in cancer patients. Other  
 CC modified p53 constructs (W13949-50, W13953-54, W13968-77) have also  
 CC been produced. Nucleic acids coding for modified p53 can be used  
 CC for cancer gene therapy.  
 SQ Sequence 363 AA;

Query Match 100.0%; Score 67; DB 1; Length 363;  
 Best Local Similarity 100.0%; Pred. No. 1.88e+00;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 26 LPENNVSPL 35  
 |||||  
 QY 1 LPENNVSPL 10

## RESULT 14

ID W13977 standard; Protein; 363 AA.  
 AC W13977;

DT 25-JUN-1997 (first entry)  
 DE Modified p53 variant p53C273R284del364-393.  
 KW p53; tumour suppressor; cancer; therapy; cell proliferation;  
 KW apoptosis; protein engineering; DNA binding.  
 OS Synthetic.  
 PN W09710843-Al.  
 PD 27-MAR-1997.  
 PF 20-SEP-1996; U15188.  
 PR 22-SEP-1995; US-004802.  
 PR 21-AUG-1996; US-697221.  
 PA (WIST-) WISTAR INST ANATOMY & BIOLOGY.  
 PI Halazonetis TD;  
 DR WPI; 97-202618/18.  
 PT R284K modified p53 protein having DNA binding ability - useful in  
 PT treatment of cancer  
 PS Example 1: 61-63; 82pp; English.  
 CC Modified p53 variant p53C273R284del364-393 (W13977) has the tumour-  
 CC derived Cys273 mutation (see also W13952), a Thr284 to Arg substn.  
 CC (see also W13949) and a deletion of the 30 C-terminal amino acids  
 CC of wild-type p53 (W13948). Cys273 is a Class I p53 tumour mutation  
 CC that affects DNA binding. The T284R substitution, introduced by  
 CC site-directed mutagenesis of p53 DNA, provides a novel p53-DNA  
 CC contact between a phosphate of the DNA backbone and p53, and  
 CC restores DNA binding. The C-terminal deletion permits in vitro  
 CC DNA binding. The construct provides the means for pharmacological  
 CC rescue of p53 function in cancer patients. Other modified p53  
 CC constructs (W13949-50, W13953-54, W13968-76) have also been  
 CC produced. Nucleic acids coding for modified p53 can be used for  
 CC cancer gene therapy.  
 SQ Sequence 363 AA;

Query Match 100.0%; Score 67; DB 1; Length 363;  
 Best Local Similarity 100.0%; Pred. No. 1.88e+00;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 26 LPENNVSPL 35  
 |||||  
 QY 1 LPENNVSPL 10

## RESULT 15

ID W13954 standard; Protein; 363 AA.  
 AC W13954;  
 DT 25-JUN-1997 (first entry)  
 DE Modified p53 variant (del364-393).  
 KW p53; tumour suppressor; cancer; therapy; cell proliferation;  
 KW apoptosis; protein engineering; DNA binding.  
 OS Synthetic.  
 PN W09710843-Al.  
 PD 27-MAR-1997.  
 PF 20-SEP-1996; U15188.  
 PR 22-SEP-1995; US-004802.  
 PR 21-AUG-1996; US-697221.  
 PA (WIST-) WISTAR INST ANATOMY & BIOLOGY.  
 PI Halazonetis TD;  
 DR WPI; 97-202618/18.  
 PT R284K modified p53 protein having DNA binding ability - useful in  
 PT treatment of cancer  
 PS Example 1: 49-51; 82pp; English.  
 CC A modified p53 variant (W13954) comprises wild-type p53 (see  
 CC also W13948) having a deletion of the C-terminal 30 amino acids,  
 CC and is obt'd. by site-directed mutagenesis of p53 DNA. Deletion of  
 CC the p53 C-terminal 30 amino acids activates the DNA binding of  
 CC common Class I p53 mutants (see also W13951-52). Novel modified  
 CC p53 variants (W13949-50, W13953-54, W13968-77), some contg.  
 CC C-terminal deletions, provide the means for pharmacological rescue  
 CC of p53 function in cancer patients. Nucleic acids coding for  
 CC modified p53 can be used for cancer gene therapy.  
 SQ Sequence 363 AA;

Query Match 100.0%; Score 67; DB 1; Length 363;  
 Best Local Similarity 100.0%; Pred. No. 1.88e+00;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 26 LPENNVLSPL 35  
| | | | | | | | | |  
QY 1 LPENNVLSPL 10

Search completed: Sat Apr 15 00:13:09 2000  
Job time : 42 secs.

\*\*\*\*\*  
M P S R E L H  
\*\*\*\*\*  
(TM)  
\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd  
MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Sat Apr 15 00:14:16 2000; MasPar time 3.09 Seconds  
Tabular output not generated. 96.683 Million cell updates/sec

Title: >US-08-452-843-13  
Description: (1-10) from US08452843.pep  
Perfect Score: 67  
Sequence: 1 LPENNVSPL 10  
Scoring table: PAM 150  
Gap 15  
Searched: 82229 seqs, 29864866 residues  
Post-processing: Minimum Match 0%  
Listing first 45 summaries  
Database: swiss-prot38  
1:swissprot  
Statistics: Mean 23.991; Variance 24.932; scale 0.962

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	67	100.0	393	1 P53_HUMAN	CELLULAR TUMOR ANTIGEN	7.71e-05
2	67	100.0	393	1 P53_CERAE	CELLULAR TUMOR ANTIGEN	7.71e-05
3	67	100.0	393	1 P53_WACMU	CELLULAR TUMOR ANTIGEN	7.71e-05
4	64	95.5	393	1 P53_WACFA	CELLULAR TUMOR ANTIGEN	5.15e-04
5	60	89.6	314	1 P53_SPEBE	CELLULAR TUMOR ANTIGEN	6.02e-03
6	52	77.6	381	1 P53_CANFA	CELLULAR TUMOR ANTIGEN	6.16e-01
7	52	77.6	386	1 P53_FELCA	CELLULAR TUMOR ANTIGEN	6.16e-01
8	51	76.1	393	1 P53_CRIGR	CELLULAR TUMOR ANTIGEN	1.06e+00
9	51	76.1	396	1 P53_MESAU	CELLULAR TUMOR ANTIGEN	1.06e+00
10	51	76.1	1752	1 RBL1_SCHPO	DNA-DIRECTED RNA POLYM	1.06e+00
11	50	74.6	785	1 DMSA_ECOLI	ANAEROBIC DIMETHYL SUL	1.83e+00
12	48	71.6	640	1 DXS_SYNY3	PROBABLE 1-DEOXYXYLULO	5.24e+00
13	47	70.1	207	1 LEXA_AERHY	LEXA REPRESSOR (EC 3.4	8.77e+00
14	47	70.1	363	1 YP62_CAEEL	PUTATIVE SERINE/THREON	8.77e+00
15	47	70.1	577	1 YLUP_ECOLI	HYPOTHETICAL 66.6 KD P	8.77e+00
16	47	70.1	1124	1 POL_FIVT2	POL POLYPROTEIN [CONTA	8.77e+00
17	46	68.7	266	1 EIA_ADEI2	EARLY EIA 29.5 KD PROT	1.45e+01
18	46	68.7	382	1 P53_SHEEP	CELLULAR TUMOR ANTIGEN	1.45e+01
19	46	68.7	386	1 P53_BOVIN	CELLULAR TUMOR ANTIGEN	1.45e+01
20	46	68.7	537	1 IL2B_RAT	INTERLEUKIN-2 RECEPTOR	1.45e+01
21	45	67.2	348	1 YLW1_CAEEL	HYPOTHETICAL 41.0 KD P	2.39e+01
22	45	67.2	640	1 PFCM_CHICK	PHOSPHOENOLPYRUVATE CA	2.39e+01
23	45	67.2	1124	1 POL_FIVSD	POL POLYPROTEIN [CONTA	2.39e+01

24	45	67.2	1124	1 POL_FIVPE	POL POLYPROTEIN [CONTA	2.39e+01
25	45	67.2	3130	1 DPOZ_HUMAN	DNA POLYMERASE ZETA CA	2.39e+01
26	44	65.7	303	1 PYP3_SCHPO	PROTEIN-TYROSINE PHOSP	3.88e+01
27	44	65.7	304	1 CAHH_VACCV	CELL SURFACE-BINDING P	3.88e+01
28	44	65.7	304	1 CAHE_VARY	CELL SURFACE-BINDING P	3.88e+01
29	44	65.7	304	1 CAHE_VACCC	CELL SURFACE-BINDING P	3.88e+01
30	44	65.7	386	1 RPA2_METVA	DNA-DIRECTED RNA POLYM	3.88e+01
31	44	65.7	388	1 YGBK_ECOLI	HYPOTHETICAL 41.3 KD P	3.88e+01
32	44	65.7	468	1 KG3H_DICDI	GLYCOCEN SYNTHASE KINA	3.88e+01
33	44	65.7	495	1 ACCD_MYCTU	PUTATIVE ACETYL-COENZY	3.88e+01
34	44	65.7	502	1 YMA0_MARPO	HYPOTHETICAL 57.7 KD P	3.88e+01
35	44	65.7	783	1 GCR_MOUSE	GLUCOCORTICOID RECEPTO	3.88e+01
36	44	65.7	795	1 GCR_RAT	GLUCOCORTICOID RECEPTO	3.88e+01
37	44	65.7	806	1 DMSA_HAEIN	ANAEROBIC DIMETHYL SUL	3.88e+01
38	44	65.7	867	1 SYA_AQUAE	ALANYL-TRNA SYNTHETASE	3.88e+01
39	43	64.2	85	1 PMRD_SALTY	POLYMYXIN B RESISTANCE	6.25e+01
40	43	64.2	275	1 RL2_BACST	50S RIBOSOMAL PROTEIN	6.25e+01
41	43	64.2	397	1 REQN_RAT	ZINC-FINGER PROTEIN NE	6.25e+01
42	43	64.2	521	1 IMAL_XENLA	IMPORTIN ALPHA-1 SUBUN	6.25e+01
43	43	64.2	621	1 RPOC_NOSCO	DNA-DIRECTED RNA POLYM	6.25e+01
44	43	64.2	712	1 CN4C_HUMAN	CAMP-DEPENDENT 3',5'-C	6.25e+01
45	43	64.2	1103	1 RPOB_CYPAA	DNA-DIRECTED RNA POLYM	6.25e+01

ALIGNMENTS

RESULT 1  
ID P53\_CERAE STANDARD; PRT; 393 AA.  
AC P13481;  
DT 01-JAN-1990 (Rel. 13, Created)  
DT 01-JAN-1990 (Rel. 13, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE CELLULAR TUMOR ANTIGEN P53.  
GN TP53  
OS Cercopithecus aethiops (Green monkey) (Grivet).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Primates; Catarrhini; Cercopithecidae; Cercopithecinae;  
OC Chlorocebus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=LIVER;  
RX MEDLINE; 90045967.  
RA RIGAUDY P., ECKHARDT W.;  
RT "Nucleotide sequence of a cDNA encoding the monkey cellular  
phosphoprotein p53.";  
RL Nucleic Acids Res. 17:8375-8375(1989).  
CC -!- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES  
GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL  
CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN  
TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION, IT IS A  
TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION  
BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF  
THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.  
APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF  
BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2  
EXPRESSION.  
CC -!- SUBCELLULAR LOCATION: NUCLEAR.  
CC -!- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY  
OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED  
IN MANY TYPES OF CANCER.  
CC -!- SIMILARITY: BELONGS TO THE P53 FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: X16384; CAA34420.1; -.  
CC PIR: S06594; S06594.  
CC HSSP: P04637; 1SAH.  
CC -----

DR PROSITE; PS00348; P53; 1.  
DR PFAM; PF00870; P53; 1.  
KW Anti-oncogene; DNA-binding; Transcription regulation; Activator;  
KW Nuclear protein; Phosphorylation; Apoptosis.  
FT DOMAIN 1 68 ASP/GLU-RICH (ACIDIC).  
FT DOMAIN 81 150 HYDROPHOBIC.  
FT DOMAIN 319 393 HIGHLY BASIC AND MAY BE INVOLVED IN  
FT DOMAIN 311 323 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
FT MOD\_RES 392 392 INTERACTION WITH DNA.  
FT MOD\_RES 392 392 PHOSPHORYLATION (BY SIMILARITY).  
SQ SEQUENCE 393 AA; 43696 MW; BBE7DC62 CRC32;  
  
Query Match 100.0%; Score 67; DB 1; Length 393;  
Best Local Similarity 100.0%; Pred. No. 7.71e-05;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Db 26 LPENNVLSP 35  
QY 1 LPENNVLSP 10  
|||||||  
|  
RESULT 2  
ID P53\_HUMAN STANDARD; PRT: 393 AA.  
AC P04637;  
DT 13-AUG-1987 (Rel. 05, Created)  
DT 01-MAR-1989 (Rel. 10, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE CELLULAR TUMOR ANTIGEN P53 (PHOSPHOPROTEIN P53).  
GN TP53.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 85230577.  
RA ZAKUT-HOURI R.; BIENZ-TADMOR B.; GIVOL D.; OREN M.;  
RT "Human p53 cellular tumor antigen: cDNA sequence and expression in  
RT COS cells.";  
RL EMBO J. 4:1251-1255(1985).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 87064416.  
RA LAMB P.; CRAWFORD L.;  
RT "Molecular cloning and in vitro expression of a cDNA clone for human  
RT cellular tumor antigen p53.";  
RL Mol. Cell. Biol. 6:1379-1385(1986).  
RN [3]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 85267676.  
RA HARLOW E.; WILLIAMSON N.M.; RALSTON R.; HELPMAN D.M.; ADAMS T.E.;  
RT "Molecular cloning and in vitro expression of a cDNA clone for human  
RT cellular tumor antigen p53.";  
RL Mol. Cell. Biol. 5:1601-1610(1985).  
RN [4]  
RP TRANSFORMED HYBRIDOMA SV-80 CELL LINE, SEQUENCE FROM N.A.  
RX MEDLINE; 87089826.  
RA HARRIS N.; BRILL E.; SHOHAT O.; PROKOCIMER M.; WOLF D.; ARAI N.;  
RA ROTTER V.;  
RT "Molecular basis for heterogeneity of the human p53 protein.";  
RL Mol. Cell. Biol. 6:4650-4656(1986).  
RN [5]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 89108008.  
RA BUCHMAN V.L.; CHUMAKOV P.M.; NINKINA N.N.; SAMARINA O.P.;  
RA GEORGIEV G.P.;  
RT "A variation in the structure of the protein-coding region of the  
RT human p53 gene.";  
RL Gene 70:245-252(1988).  
RN [6]  
RP SEQUENCE OF 101-393 FROM N.A.  
RX MEDLINE; 85126934.  
RA MATIASHEWSKI G.; LAMB P.; PIM D.; PEACOCK J.; CRAWFORD L.;  
RA BENCHIMOL S.;  
RT "Isolation and characterization of a human p53 cDNA clone: expression

of the human p53 gene.";  
RL EMBO J. 3:3257-3262(1984).  
RN [7]  
RP NUCLEAR LOCALIZATION SIGNAL.  
RX MEDLINE; 90191730.  
RA ADDISON C.; JENKINS J.R.; STURZBECHER H.-W.;  
RT "The p53 nuclear localisation signal is structurally linked to a  
RT p34cdc2 kinase motif.";  
RL Oncogene 5:423-426(1990).  
RN [8]  
RP PHOSPHORYLATION BY P60/CDC2 AND CYCLIN B/CDC2.  
RX MEDLINE; 90280456.  
RA BISCHOFF J.R.; FRIEDMAN P.N.; MARSHAK D.R.; PRIVES C.; BEACH D.;  
RT "Human p53 is phosphorylated by p60-cdc2 and cyclin B-cdc2.";  
RL Proc. Natl. Acad. Sci. U.S.A. 87:4766-4770(1990).  
RN [9]  
RP DEPHOSPHORYLATION BY PP2A.  
RX MEDLINE; 91172186.  
RA SCHEIDTMANN K.H.; MUMBY M.C.; RUNDELL K.; WALTER G.;  
RT "Dephosphorylation of simian virus 40 large-T antigen and p53 protein  
RT by protein phosphatase 2A: inhibition by small-t antigen.";  
RL Mol. Cell. Biol. 11:1996-2003(1991).  
RN [10]  
RP STRUCTURE BY NMR OF 319-360.  
RX MEDLINE; 94294808.  
RA CLORE G.M.; OMICHINSKI J.G.; SAKAGUCHI K.; ZAMBRANO N.; SAKAMOTO H.;  
RA APPELLA E.; GRONENBORN A.M.;  
RT "High-resolution structure of the oligomerization domain of p53 by  
RT multidimensional NMR.";  
RL Science 265:386-391(1994).  
RN [11]  
RP STRUCTURE BY NMR OF 325-355.  
RX MEDLINE; 95292092.  
RA LEE W.; HARVEY T.S.; YIN Y.; YAU P.; LITCHEFIELD D.; ARROWSMITH C.H.;  
RT "Solution structure of the tetrameric minimum transforming domain of  
RT p53.";  
RL Nat. Struct. Biol. 1:877-890(1994).  
RN [12]  
RP STRUCTURE BY NMR OF 326-354.  
RX MEDLINE; 98026899.  
RA MCCOY M.; STAVRIDIS E.S.; WATERMAN J.L.; WIECZOREK A.M.; OPELLA S.J.;  
RA HALAZONETIS T.D.;  
RT "Hydrophobic side-chain size is a determinant of the  
RT three-dimensional structure of the p53 oligomerization domain.";  
RL EMBO J. 16:6230-6236(1997).  
RN [13]  
RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 94-289.  
RX MEDLINE; 94294806.  
RA CHO Y.; GORINA S.; JEFFREY P.D.; PAVLETICH N.P.;  
RT "Crystal structure of a p53 tumor suppressor-DNA complex:  
RT understanding tumorigenic mutations.";  
RL Science 265:346-355(1994).  
RN [14]  
RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS) OF 13-29 IN COMPLEX WITH MDM2.  
RX MEDLINE; 97081050.  
RA KUSSIE P.H.; GORINA S.; MARECHAL V.; ELENBAAS B.; MOREAU J.;  
RA LEVINE A.J.; PAVLETICH N.P.;  
RT "Structure of the MDM2 oncoprotein bound to the p53 tumor suppressor  
RT transactivation domain.";  
RL Science 274:948-953(1996).  
RN [15]  
RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 97-287 IN COMPLEX WITH 53BP2.  
RX MEDLINE; 97035414.  
RA GORINA S.; PAVLETICH N.P.;  
RT "Structure of the p53 tumor suppressor bound to the ankyrin and SH3  
RT domains of 53BP2.";  
RL Science 274:1001-1005(1996).  
RN [16]  
RP REVIEW.  
RX MEDLINE; 94090335.  
RA HARRIS C.C.;  
RT "p53: at the crossroads of molecular carcinogenesis and risk  
RT assessment."



RL Science 262:1980-1981(1993).  
RN [17]  
RP REVIEW ON VARIANTS.  
RX MEDLINE: 91289156.  
RA HOOLSTEIN M., SIDRANSKY D., VOGELSTEIN B., HARRIS C.C.;  
RT "p53 mutations in human cancers."  
RL Science 253:49-53(1991).  
RN [18]  
RP REVIEW ON VARIANTS.  
RX MEDLINE: 96271983.  
RA DE VRIES E.M.G., RICKE D.O., DE VRIES T.N., HARTMANN A., BLASZYK H.,  
RA LIAO D., SOUSSI T., KOVACH J.S., SOMMER S.S.;  
RT "Database of mutations in the p53 and APC tumor suppressor genes  
designed to facilitate molecular epidemiological analyses."  
RL Hum. Mutat. 7:202-213(1996).  
RN [19]  
RP VARIANTS ARG-72.  
RX MEDLINE: 91153807.  
RA OLSCHWANG S., LAURENT-PUIG P., VASSAL A., SALMON R.-J., THOMAS G.;  
RT "Characterization of a frequent polymorphism in the coding sequence  
of the p53 gene in colonic cancer patients and a control  
population."  
RL Hum. Genet. 86:369-370(1991).  
RN [20]  
RP VARIANTS LFS THR-133.  
RX MEDLINE: 92034774.  
RA LAW J.C., STRONG L.C., CHIDAMBARAM A., FERRELL R.E.;  
RT "A germ line mutation in exon 5 of the p53 gene in an extended cancer  
family."  
RL Cancer Res. 51:6385-6387(1991).  
RN [21]  
RP VARIANTS LFS CYS-245; TRP-248; PRO-252 AND LYS-258.  
RX MEDLINE: 91057657.  
RA MALKIN D., LI F.P., STRONG L.C., FRAUMENI J.F. JR., NELSON C.E.,  
RA KIM D.H., KASSEL J., GRYKA M.A., BISCHOFF F.Z., TAINSKY M.A.,  
RA FRIEND S.H.;  
RT "Germ line p53 mutations in a familial syndrome of breast cancer,  
sarcomas, and other neoplasms."  
RL Science 250:1233-1238(1990).  
RN [22]  
RP VARIANTS LFS ASP-245.  
RX MEDLINE: 91080929.  
RA SRIVASTAVA S., ZOU Z., PIROLLO K., BLATTNER W., CHANG E.H.;  
RT "Germ-line transmission of a mutated p53 gene in a cancer-prone  
family with Li-Fraumeni syndrome."  
RL Nature 348:747-749(1990).  
RN [23]  
RP VARIANTS LFS LEU-272.  
RX MEDLINE: 92147883.  
RA FELIX C.A., NAU M.M., TAKAHASHI T., MITSUDOMI T., CHIBA I.,  
RA POPLACK D.G., REAMAN G.H., COLE D.E., LETTERIO J.J., WHANG-PENG J.,  
RA KNUDSEN T., MINNA J.D.;  
RT "Hereditary and acquired p53 gene mutations in childhood acute  
lymphoblastic leukemia."  
RL J. Clin. Invest. 89:640-647(1992).  
RN [24]  
RP VARIANTS LFS HIS-273 AND VAL-325.  
RX MEDLINE: 92228023.  
RA MALKIN D., JOLLY K.W., BARBIER N., LOOK A.T., FRIEND S.H.,  
RA GEBHART M.C., ANDERSEN T.I., BORRESEN A.-L., LI F.P., GARBER J.,  
RA STRONG L.C.;  
RT "Germline mutations of the p53 tumor-suppressor gene in children and  
RT young adults with second malignant neoplasms."  
RL New Engl. J. Med. 326:1309-1315(1992).  
RN [25]  
RP VARIANTS BREAST TUMORS GLN-132; SER-249; LYS-280 AND LYS-285.  
RX MEDLINE: 90295284.  
RA BARTEK J., IGGO R., GANNON J., LANE D.P.;  
RT "Genetic and immunochemical analysis of mutant p53 in human breast  
RT cancer cell lines."  
RL Oncogene 5:893-899(1990).  
RN [26]  
RP VARIANTS COLON TUMORS PHE-241 AND HIS-273.  
RX MEDLINE: 91017544.  
RA RODRIGUES N.R., ROWAN A., SMITH M.E.F., KERR I.B., BODMER W.F.,  
RA GANNON J.V., LANE D.P.;  
RT "Note: remainder of annotations omitted."  
RL Query Match 100.0%; Score 67; DB 1; Length 393;  
Best Local Similarity 100.0%; Pred. No. 7.71e-05;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Db 26 LPENNVLSP 35  
QY 1 LPENNVLSP 10  
RESULT 3  
ID P53 MACMU STANDARD; PRT; 393 AA.  
AC P58424;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE CELLULAR TUMOR ANTIGEN P53.  
GN TP53 OR P53.  
OS Macaca mulatta (Rhesus macaque).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Primates; Catarrhini; Cercopithecoidea; Cercopithecinae;  
OC Macaca.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA KHAN M.A., HANSEN C., WELSH J.A., BENNETT W.P.;  
RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES  
CC GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL  
CC CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN  
CC TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION. IT IS A  
CC TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION  
CC BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF  
CC THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.  
CC APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF  
CC BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2  
CC EXPRESSION.  
CC -!- SUBCELLULAR LOCATION: NUCLEAR.  
CC -!- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY  
CC OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED  
CC IN MANY TYPES OF CANCER.  
CC -!- SIMILARITY: BELONGS TO THE P53 FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; U48956; AAB91534.1; -.  
DR HSSP; P04637; ISAH.  
DR PROSITE; PS00348; P53; 1.  
DR PFAM; PF00870; P53; 1.  
KW Anti-oncogene; DNA-binding; Transcription regulation; Activator;  
FT Nucleic acid binding site; Phosphorylation; Apoptosis.  
KW DOMAIN 1 80 ASP/GLU-RICH (ACIDIC).  
FT DOMAIN 81 150 HYDROPHOBIC.  
FT DOMAIN 319 393 HIGHLY BASIC AND MAY BE INVOLVED IN  
FT INTERACTION WITH DNA  
FT DOMAIN 311 323 NUCLEAR LOCALIZATION SIGNAL.  
FT MOD\_RES 392 392 PHOSPHORYLATION (BY SIMILARITY).  
SQ SEQUENCE 393 AA; 43655 MW; 11A9B7F8 CRC32;  
Query Match 100.0%; Score 67; DB 1; Length 393;  
Best Local Similarity 100.0%; Pred. No. 7.71e-05;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Db 26 LPENNVLSP 35
   |||||
QY 1 LPENNVLSP 10

RESULT 4
ID P53_MAFCA STANDARD; PRT; 393 AA.
AC P56423;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DE 15-JUL-1998 (Rel. 36, Last annotation update)
DE CELLULAR TUMOR ANTIGEN P53.
GN TP53 OR P53.
OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Cercopithecoidea; Cercopithecinae;
OC Macaca.
RN [1]
RP SEQUENCE FROM N.A.
RA KHAN M.A., HANSEN C., WELSH J.A., BENNETT W.P.;
RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES
CC GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL
CC CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN
CC TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION, IT IS A
CC TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION
CC BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF
CC THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.
CC APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF
CC BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2
CC EXPRESSION.
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY
CC OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED
CC IN MANY TYPES OF CANCER.
CC -1- SIMILARITY: BELONGS TO THE P53 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U48957; AAB91535.1;
DR HSP; P04637; 1SAH.
DR PROSITE; PS00348; P53; 1.
DR PFAM; PF00870; P53; 1.
KW Anti-oncogene; DNA-binding; Transcription regulation; Activator;
KW Nuclear protein; Phosphorylation; Apoptosis.
FT DOMAIN 1 80
FT DOMAIN 81 150
FT DOMAIN 151 393
FT HIGHLY BASIC AND MAY BE INVOLVED IN
FT INTERACTION WITH DNA.
FT DOMAIN 311 323
FT NUCLEAR LOCALIZATION SIGNAL.
FT MOD RES 392 392
FT PHOSPHORYLATION (BY SIMILARITY).
SQ SEQUENCE 393 AA; 43678 MW; 2499AC47 CRC32;

Query Match 95.5%; Score 64; DB 1; Length 393;
Best Local Similarity 90.0%; Pred. No. 5.15e-04;
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 26 LPENNVLSP 35
   |||||
QY 1 LPENNVLSP 10

RESULT 5
ID P53_SPEBE STANDARD; PRT; 314 AA.
AC Q84662;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)

Query Match 89.6%; Score 60; DB 1; Length 314;
Best Local Similarity 90.0%; Pred. No. 6.02e-03;
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 6 LPENNVLSPV 15
   |||||
QY 1 LPENNVLSP 10

RESULT 6
ID P53_CANFA STANDARD; PRT; 381 AA.
AC Q29537;
DT 01-NOV-1997 (Rel. 35, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE CELLULAR TUMOR ANTIGEN P53.
GN TP53 OR P53.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Carnivora; Fissipedia; Canidae; Canis.
RN [1]
RP SEQUENCE FROM N.A.
RP TISSUE=LEUKOCYTE;
```

```

RX MEDLINE: 98178696.
RA VELDHOEN N., MILNER J.;
RT "Isolation of canine p53 cDNA and detailed characterization of the
RT full length canine p53 protein.";
RL Oncogene 16:1077-1084(1998).
RN [2]
RP SEQUENCE OF 25-300 FROM N.A.
RC STRAIN-BEAGLE;
RX MEDLINE: 95323915.
RA KRAGEL S.A., PAZZI K.A., MADEWELL B.R.;
RT "Sequence analysis of canine p53 in the region of exons 3-8.";
RL Cancer Lett. 92:181-186(1995).
CC -!- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES
CC GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL
CC CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN
CC TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION, IT IS A
CC TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION
CC BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF
CC THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.
CC APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF
CC BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2
CC EXPRESSION.
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY
CC OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED
CC IN MANY TYPES OF CANCER.
CC -!- SIMILARITY: BELONGS TO THE P53 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF060514; AAC16909.1; -.
DR EMBL: S77819; AAB42022.1; -.
DR HSSP: P04637; 1YCS.
DR PROSITE: PS00348; P53; 1.
DR PFAM: PF00870; P53; 1.
KW Anti-oncogene; DNA-binding; Transcription regulation; Activator;
KW Nuclear protein; Phosphorylation; Apoptosis.
FT DOMAIN 1 59 ASP/GLU-RICH (ACIDIC).
FT DOMAIN 307 381 HYDROPHOBIC.
FT DOMAIN 299 311 INTERACTION WITH DNA (BY SIMILARITY).
FT DOMAIN 380 380 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT MOD_RES 68 137 PHOSPHORYLATION (BY SIMILARITY).
SQ SEQUENCE 381 AA; 42486 MW; 70210B63 CRC32;

Query Match 77.6%; Score 52; DB 1; Length 381;
Best Local Similarity 88.9%; Pred. No. 6.16e-01;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 26 LPENNVLS 34
| | | | | | |
QY 1 LPENNVLS P 9

RESULT 7
ID P53_FELCA STANDARD; PRT; 386 AA.
AC P41685;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE CELLULAR TUMOR ANTIGEN P53.
GN TP53.
OS Felis silvestris catus (Cat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Carnivora; Fissipedia; Felidae; Felis.
RN [1]
RP SEQUENCE FROM N.A.

```

```

RC TISSUE-LYMPH NODE;
RX MEDLINE: 94333960.
RA OKUDA M., UMEDA A., SAKAI T., OHASHI T., MOMOI Y., YOUN H.Y.,
RA WATARI T., GOITSUKA R., TSUJIMOTO H., HASEGAWA A.;
RT "Cloning of feline p53 tumor-suppressor gene and its aberration in
RT hematopoietic tumors.";
RL Int. J. Cancer 58:602-607(1994).
RN [2]
RP SEQUENCE OF 34-354 FROM N.A.
RX MEDLINE: 94114699.
RA OKUDA M., UMEDA A., MATSUMOTO Y., MOMOI Y., WATARI T., GOITSUKA R.,
RA O'BRIEN S.J., TSUJIMOTO H., HASEGAWA A.;
RT "Molecular cloning and chromosomal mapping of feline p53 tumor
RT suppressor gene.";
RL J. Vet. Med. Sci. 55:801-805(1993).
CC -!- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES
CC GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL
CC CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN
CC TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION, IT IS A
CC TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION
CC BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF
CC THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.
CC APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF
CC BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2
CC EXPRESSION.
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY
CC OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED
CC IN MANY TYPES OF CANCER.
CC -!- SIMILARITY: BELONGS TO THE P53 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D26608; BAA05653.1; -.
DR EMBL: D16460; BAA03927.1; -.
DR HSSP: P04637; 1SNH.
DR PROSITE: PS00348; P53; 1.
DR PFAM: PF00870; P53; 1.
KW Anti-oncogene; DNA-binding; Transcription regulation; Activator;
KW Nuclear protein; Phosphorylation; Apoptosis.
FT DOMAIN 1 59 ASP/GLU-RICH (ACIDIC).
FT DOMAIN 304 316 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT MOD_RES 385 385 PHOSPHORYLATION (BY SIMILARITY).
FT CONFLICT 285 285 K -> R (IN REF. 2).
SQ SEQUENCE 386 AA; 42692 MW; D6C7132A CRC32;

Query Match 77.6%; Score 52; DB 1; Length 386;
Best Local Similarity 88.9%; Pred. No. 6.16e-01;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 26 LPENNVLS 34
| | | | | | |
QY 1 LPENNVLS P 9

RESULT 8
ID P53_CRIGR STANDARD; PRT; 393 AA.
AC O09185; O64397; P97258; P97788;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE CELLULAR TUMOR ANTIGEN P53.
GN TP53 OR P53.
OS Cricetulus griseus (Chinese hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae; Cricetulus.
RN [1]

```

```
RP SEQUENCE FROM N.A.
RA CHAUNG W., MI L.J., BOORSTEIN R.J.;
RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RN SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RX MEDLINE; 97183659.
RA LEE H., LARNER J.M., HAMLIN J.L.;
RL "Cloning and characterization of Chinese hamster p53 cDNA.";
RL Gene 184:177-183(1997).
RN [3]
RN SEQUENCE FROM N.A.
RP TISSUE-EMBRYONIC FIBROBLAST;
RA SHIMIZU T., NIKAIKO O., SUZUKI F.;
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
RN [1]
RN FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES
CC GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL
CC CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN
CC TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION, IT IS A
CC TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION
CC BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF
CC THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.
CC APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF
CC BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2
CC EXPRESSION.
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY
CC OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED
CC IN MANY TYPES OF CANCER.
CC -!- SIMILARITY: BELONGS TO THE P53 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Y08900; CAA70108.1; -
DR EMBL; Y08901; CAA70109.1; -
DR EMBL; U50395; AAC53040.1; -
DR EMBL; D86070; BAA13004.1; -
DR HSSP; P04637; LYCQ.
DR PROSITE; PS00348; P53; 1.
DR PFAM; PF00870; P53; 1.
KW Anti-oncogene; DNA-binding; Transcription regulation; Activator;
KW Nuclear protein; Phosphorylation; Apoptosis.
FT DOMAIN 1 74 ASP/GLU-RICH (ACIDIC).
FT DOMAIN 75 150 HYDROPHOBIC.
FT DOMAIN 316 390 HIGHLY BASIC AND MAY BE INVOLVED IN
FT FT INTERACTION WITH DNA.
FT FT NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT FT PHOSPHORYLATION (BY SIMILARITY).
FT MOD_RES 392 392 L -> Q (IN CELL LINE V79-4).
FT VARIANT 133 133 C -> W (IN CELL LINE V79-4).
FT VARIANT 135 135 Y -> F (IN REF. 2).
FT CONFLICT 103 103 Y -> F (IN REF. 2).
SQ SEQUENCE 393 AA; 43378 MW; 402EB149 CRC32;

Query Match 76.1%; Score 51; DB 1; Length 393;
Best Local Similarity 80.0%; Pred. No. 1.06e+00;
Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 26 LPPNNVLSL 35
QY 1 LPPNNVLSL 10

RESULT 9
ID P53_MESAU STANDARD; PRT; 396 AA.
AC Q00366; P97276;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)

RP SEQUENCE FROM N.A.
RA CHAUNG W., MI L.J., BOORSTEIN R.J.;
RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RN SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RX MEDLINE; 97183659.
RA LEE H., LARNER J.M., HAMLIN J.L.;
RL "Cloning and characterization of Chinese hamster p53 cDNA.";
RL Gene 184:177-183(1997).
RN [3]
RN SEQUENCE FROM N.A.
RP TISSUE-EMBRYONIC FIBROBLAST;
RA SHIMIZU T., NIKAIKO O., SUZUKI F.;
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
RN [1]
RN FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES
CC GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL
CC CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN
CC TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION, IT IS A
CC TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION
CC BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF
CC THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.
CC APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF
CC BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2
CC EXPRESSION.
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY
CC OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED
CC IN MANY TYPES OF CANCER.
CC -!- SIMILARITY: BELONGS TO THE P53 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M75144; AAA37085.1; -
DR EMBL; U07182; AAB41344.1; -
DR PIR; JH0633; JH0633.
DR HSSP; P04637; LYCQ.
DR PROSITE; PS00348; P53; 1.
DR PFAM; PF00870; P53; 1.
KW Anti-oncogene; DNA-binding; Transcription regulation; Activator;
KW Nuclear protein; Phosphorylation; Apoptosis.
FT DOMAIN 1 77 ASP/GLU-RICH (ACIDIC).
FT DOMAIN 78 153 HYDROPHOBIC.
FT DOMAIN 319 393 HIGHLY BASIC AND MAY BE INVOLVED IN
FT FT INTERACTION WITH DNA.
FT FT NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT MOD_RES 395 395 PHOSPHORYLATION (BY SIMILARITY).
FT CONFLICT 188 188 G -> S (IN REF. 2).
SQ SEQUENCE 396 AA; 43631 MW; C2668ADE CRC32;

Query Match 76.1%; Score 51; DB 1; Length 396;
Best Local Similarity 80.0%; Pred. No. 1.06e+00;
Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 26 LPPNNVLSL 35
QY 1 LPPNNVLSL 10

RESULT 10
ID RPBL_SCHPO STANDARD; PRT; 1752 AA.
AC P36594;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
```

DE DNA-DIRECTED RNA POLYMERASE II LARGEST SUBUNIT (EC 2.7.7.6)  
 DE (RNA POLYMERASE II SUBUNIT 1).  
 GN RPB1.  
 OS Schizosaccharomyces pombe (fission yeast).  
 OC Eukaryota; Fungi; Ascomycota; Archiascomycetes;  
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
 OC Schizosaccharomyces.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=972;  
 RX MEDLINE; 91187661.  
 RA AZUMA Y., YARNAGISHI M., UESHIMA R., ISHIHAMA A.;  
 RT "Cloning and sequencing determination of the Schizosaccharomyces pombe  
 RT rpb1 gene encoding the largest subunit of RNA polymerase II.";  
 RL Nucleic Acids Res. 19:461-468(1991).  
 CC -1- FUNCTION: RNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION  
 CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS  
 CC SUBSTRATES.  
 CC -1- CATALYTIC ACTIVITY: N NUCLEOSIDE TRIPHOSPHATE - N PYROPHOSPHATE +  
 CC RNA(N).  
 CC -1- SUBUNIT: RNA POLYMERASE II CONSISTS OF 10 DIFFERENT SUBUNITS.  
 CC THIS SUBUNIT IS THE LARGEST COMPONENT OF RNA POLYMERASE II.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -1- PTM: THE TANDEM 7 RESIDUES REPEATS CAN BE HIGHLY PHOSPHORYLATED.  
 CC THE PHOSPHORYLATION ACTIVATES POL2.  
 CC -1- MISCELLANEOUS: THREE DISTINCT ZINC-CONTAINING RNA POLYMERASES ARE  
 CC FOUND IN EUKARYOTIC NUCLEI: POLYMERASE I FOR THE RIBOSOMAL RNA  
 CC PRECURSOR, POLYMERASE II FOR THE MRNA PRECURSOR, AND POLYMERASE  
 CC III FOR 5S AND TRNA GENES.  
 CC -1- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA' CHAIN FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL; X56564; CAA39916.1; -.  
 DR PIR; S26849; S26849.  
 DR PROSITE; PS00115; RNA\_POL\_II\_REPEAT; 24.  
 DR PFAM; PF00623; RNA\_pol\_A; 1.  
 KW Transferase; DNA-directed RNA polymerase; Transcription; Zinc; Repeat;  
 KW DNA-binding; Nuclear protein; Phosphorylation; Zinc-finger.  
 FT 2N.FING 69 85 C2H2-TYPE (POTENTIAL).  
 FT DOMAIN 1554 1752 CARBOXYL-TERMINAL 7-RESIDUE REPEATS.  
 SQ SEQUENCE 1752 AA; 194161 MW; B7CFE872 CRC32;  
 Query Match 76.1%; Score 51; DB 1; Length 1752;  
 Best Local Similarity 70.0%; Pred. No. 1.06e+00;  
 Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
 Db 202 LPKRLLSPL 211  
 |||: |||||  
 QY 1 LPENNLSPL 10

MEDLINE; 89096500.  
 RA BILIOUS P.T., COLE S.T., ANDERSON W.F., WEINER J.H.;  
 RT "Nucleotide sequence of the dmsABC operon encoding the anaerobic  
 RT dimethylsulphoxide reductase of Escherichia coli.";  
 RL Mol. Microbiol. 2:785-795(1988).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12 / MG1655;  
 RX MEDLINE; 97426617.  
 RA BLATTNER F.R., PLUNKETT G. III, BLOCH C.A., PERNA N.T., BURLAND V.,  
 RA RILEY M., COLLADO-VIDES J., GLASNER F.D., RODE C.K., MAYHEW G.F.,  
 RA GREGOR J., DAVIS N.W., KIRKPATRICK H.A., GORDEN M.A., ROSE D.J.,  
 RA MAU B., SHAO Y.;  
 RT "The complete genome sequence of Escherichia coli K-12.";  
 RL Science 277:1453-1474(1997).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12.  
 RX MEDLINE; 97061202.  
 RA OSHIMA T., AIBA H., BABA T., FUJITA K., HAYASHI K., HONJO A.,  
 RA IKEMOTO K., INADA T., ITOH T., KAJIHARA M., KANAI K., KASHIMOTO K.,  
 RA KIMURA S., KITAGAWA M., MAKINO K., MASUDA S., MIKI T., MIZOBUCHI K.,  
 RA MORI H., MOTOMURA K., NAKAMURA Y., NASHIMOTO H., NISHIO Y., SAITO N.,  
 RA SAMPEI G., SEKI Y., TAGAMI H., TAKEMOTO K., WADA C., YAMANOTO Y.,  
 RA YANO M., HORIOUCHI T.;  
 RT "A 718-Kb DNA sequence of the Escherichia coli K-12 genome  
 RT corresponding to the 12.7-28.0 min region on the linkage map.";  
 RL DNA Res. 3:137-155(1996).  
 RN [4]  
 RP MUTAGENESIS.  
 RX MEDLINE; 94171715.  
 RA TRIEBER C.A., ROTHERY R.A., WEINER J.H.;  
 RT "Multiple pathways of electron transfer in dimethyl sulfoxide  
 RT reductase of Escherichia coli.";  
 RL J. Biol. Chem. 269:7103-7109(1994).  
 CC -1- FUNCTION: TERMINAL REDUCTASE DURING ANAEROBIC GROWTH ON  
 CC VARIOUS SULFOXIDE AND N-OXIDE COMPOUNDS. ALLOWS E.COLI TO GROW  
 CC ANAEROBICALLY ON ME(2)SO AS RESPIRATORY OXIDANT.  
 CC -1- CATALYTIC ACTIVITY: REDUCES VARIOUS N-OXIDE AND SULFOXIDE  
 CC COMPOUNDS INCLUDING TRIMETHYLAMINE N-OXIDE.  
 CC -1- COFACTOR: MOLYBDENUM (MOLYBDOPTERIN); MAY BIND A 4FE-4S CLUSTER.  
 CC -1- SUBUNIT: HOMODIMER. THE COMPLEX CONSIST OF THREE SUBUNITS: DMSA,  
 CC THE REDUCTASE; DMSB, AN ELECTRON TRANSFER PROTEIN, AND DMSC,  
 CC A MEMBRANE ANCHOR PROTEIN.  
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC FACE OF THE MEMBRANE.  
 CC -1- SIMILARITY: BELONGS TO THE PROKARYOTIC MOLYBDOPTERIN-CONTAINING  
 CC OXIDOREDUCTASE FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL; J03412; AAA83843.1; -.  
 DR EMBL; AE000191; AAC73980.1; -.  
 DR EMBL; D90727; BAA35626.1; -.  
 DR PIR; S03785; S03785.  
 DR HSSP; Q57366; ICXT.  
 DR ECGENE; EG10232; DMSA.  
 DR PROSITE; PS00551; MOLYBDOPTERIN\_PROK\_1; 1.  
 DR PROSITE; PS00490; MOLYBDOPTERIN\_PROK\_2; 1.  
 DR PROSITE; PS00932; MOLYBDOPTERIN\_PROK\_3; 1.  
 DR PFAM; PF00384; molybdopterin; 1.  
 DR PFAM; PF01568; Molybdop\_binding; 1.  
 KW Oxidoreductase; Signal; Molybdenum; 4Fe-4S; Iron-sulfur.  
 FT SIGNAL 1 16  
 FT CHAIN 17 785 ANAEROBIC DIMETHYL SULFOXIDE REDUCTASE  
 FT CHAIN A.  
 FT METAL 34 34 IRON-SULFUR (4FE-4S) (BY SIMILARITY).  
 FT METAL 38 38 IRON-SULFUR (4FE-4S) (BY SIMILARITY).

FT METAL 42 42 IRON-SULFUR (4FE-4S) (BY SIMILARITY).  
FT METAL 75 75 IRON-SULFUR (4FE-4S) (BY SIMILARITY).  
SQ SEQUENCE 785 AA; 87449 MW; 9C1ADEB0 CRC32;  
Query Match 74.6%; Score 50; DB 1; Length 785;  
Best Local Similarity 60.0%; Pred. No. 1.83e+00;  
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
Db 637 LPEGVDIPL 646  
| | | | | | | | | |  
QY 1 LPENNVLSP 10  
RESULT 12  
ID DXS.SYN3 STANDARD; PRT; 640 AA.  
AC P73067;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 15-JUL-1998 (Rel. 36, Last annotation update)  
DE PROBABLE 1-DEOXYXYLOSE-5-PHOSPHATE SYNTHASE (DXP SYNTHASE).  
GN DXS OR SL11945.  
OS Synechocystis sp. (strain PCC 6803).  
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 97061201.  
RA KANEKO T., SATO S., KOTANI H., TANAKA A., ASAMIZU E., NAKAMURA Y.,  
RA MIYAJIMA N., HIROSAWA M., SUGTURA M., SASAMOTO S., KIMURA T.,  
RA HOSOUCHI T., MATSUNO A., MURAKI A., NAKAZAKI N., NARUO K.,  
RA OKUMURA S., SHIMPO S., TAKEUCHI C., WADA T., WATANABE A.,  
RA YAMADA M., YASUDA M., TABAYA S.;  
RT "Sequence analysis of the genome of the unicellular cyanobacterium  
RT Synechocystis sp. strain PCC6803. II. Sequence determination of the  
RT entire genome and assignment of potential protein-coding regions.";  
RL DNA Res. 3:109-136(1996).  
CC -!- FUNCTION: CATALYZES THE ACYLON CONDENSATION REACTION BETWEEN C  
CC ATOMS 2 AND 3 OF PYRUVATE AND GLYCERALDEHYDE 3-PHOSPHATE TO YIELD  
CC 1-DEOXY-D-XYLOSE-5-PHOSPHATE (DXP) (BY SIMILARITY).  
CC -!- COPACITOR: THIAMINE PYROPHOSPHATE (BY SIMILARITY).  
CC -!- PATHWAY: IN THE BIOSYNTHETIC PATHWAY TO ISOPRENOIDS, THIAMINE, AND  
CC PYRIDOXOL (BY SIMILARITY).  
CC -!- SIMILARITY: BELONGS TO THE TRANSKETOLASE FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; D90903; BAA17089.1; -;  
DR PROSITE; PS00801; TRANKETOLASE\_1; 1.  
DR PROSITE; PS00802; TRANKETOLASE\_2; 1.  
KW Flavoprotein; Thiamine pyrophosphate; Isoprene biosynthesis;  
KW Thiamine biosynthesis.  
SQ SEQUENCE 640 AA; 69328 MW; EED14440 CRC32;  
Query Match 71.6%; Score 48; DB 1; Length 640;  
Best Local Similarity 70.0%; Pred. No. 5.24e+00;  
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
Db 580 LMDNNVLSP 589  
| | | | | | | | | |  
QY 1 LPENNVLSP 10  
RESULT 13  
ID LEXA.AERHY STANDARD; PRT; 207 AA.  
AC Q44069;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)

DE LEXA REPRESSOR (EC 3.4.21.88).  
GN LEXA.  
OS Aeromonas hydrophila.  
OC Bacteria; Proteobacteria; gamma subdivision; Aeromonas group;  
OC Aeromonas.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-ATCC 7966;  
RX MEDLINE; 95172406.  
RA RIERA J., BARBE J.;  
RT "Cloning, sequence and regulation of expression of the lexA gene of  
RT Aeromonas hydrophila.";  
RL Gene 154:71-75(1995).  
CC -!- FUNCTION: REPRESSSES A NUMBER OF GENES INVOLVED IN THE RESPONSE TO  
CC DNA DAMAGE (SOS RESPONSE), INCLUDING RECA AND LEXA. BINDS TO A  
CC 16 BP PALINDROMIC SEQUENCE. IN THE PRESENCE OF SINGLE-STRANDED  
CC DNA, RECA INTERACTS WITH LEXA CAUSING AN AUTOCATALYTIC CLEAVAGE  
CC WHICH DISRUPTS THE DNA-BINDING PART OF LEXA, LEADING TO  
CC DEREGULATION OF THE SOS REGULATION AND EVENTUALLY DNA REPAIR  
CC (BY SIMILARITY).  
CC -!- CATALYTIC ACTIVITY: HYDROLYSIS OF 84-ALA-|-GLY-85 BOND IN  
CC REPRESSOR LEXA.  
CC -!- SUBUNIT: HOMODIMER (BY SIMILARITY).  
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S24 ALSO KNOWN AS THE  
CC UMUD/LEXA FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; X77263; CAA54479.1; -;  
DR HSPF; P03033; 1LEB.  
DR PFAM; PF00717; Peptidase\_S24; 1.  
KW Transcription regulation; Repressor; DNA damage;  
KW Autocatalytic cleavage; Hydrolyase; DNA replication; SOS response;  
KW DNA-binding. 28 48 H-T-H MOTIF (BY SIMILARITY).  
FT SITE 89 90 CLEAVAGE (AUTO-) (BY SIMILARITY).  
FT ACT\_SITE 123 123 INVOLVED IN AUTO-CLEAVAGE  
FT ACT\_SITE 161 161 INVOLVED IN AUTO-CLEAVAGE  
FT ACT\_SITE 161 161 INVOLVED IN AUTO-CLEAVAGE  
FT ACT\_SITE 161 161 INVOLVED IN AUTO-CLEAVAGE  
FT ACT\_SITE 161 161 INVOLVED IN AUTO-CLEAVAGE  
SQ SEQUENCE 207 AA; 22898 MW; A7AA3AB3 CRC32;  
Query Match 70.1%; Score 47; DB 1; Length 207;  
Best Local Similarity 70.0%; Pred. No. 8.77e+00;  
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
Db 173 LPENEELSPI 182  
| | | | | | | | | |  
QY 1 LPENNVLSP 10  
RESULT 14  
ID YR62.CAEEL STANDARD; PRT; 363 AA.  
AC Q20347; C01707;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 15-DEC-1999 (Rel. 39, Last annotation update)  
DE PUTATIVE SERINE/THREONINE-PROTEIN KINASE F42G10.2 IN CHROMOSOME II  
DE (EC 2.7.1.-).  
GN F42G10.2  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;  
OC Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RA HARRIS B., LENNARD N.;

RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.  
CC -!- SIMILARITY: WITH THE CONSERVED CATALYTIC DOMAINS OF SER/THR-  
CC PROTEIN KINASES. BELONGS TO THE MAP KINASE KINASE FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; Z95122; CAB08355.1; -  
DR EMBL; Z48230; CAB08355.1; JOINED.  
DR EMBL; Z48230; CAB08355.1; -  
DR EMBL; Z95122; CAB08355.1; JOINED.  
DR HSP; P24941; LHCK.  
DR WORMPEP; P42G10.2; CE10328.  
DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
DR PROSITE; PS00108; PROTEIN\_KINASE\_ST; 1.  
DR PROSITE; PS00111; PROTEIN\_KINASE\_DOM; 1.  
DR PFAM; PF00069; PKINASE; 1.  
KW Hypothetical protein; Transferase; Serine/threonine-protein kinase;  
KW ATP-binding.  
FT DOMAIN 66 330 PROTEIN KINASE.  
FT NP\_BIND 72 80 ATP (BY SIMILARITY).  
FT BINDING 95 95 ATP (BY SIMILARITY).  
FT ACT\_SITE 194 194 BY SIMILARITY.  
SQ SEQUENCE 363 AA; 41197 MW; BAC61B3E CRC32;  
  
Query Match 70.1%; Score 47; DB 1; Length 363;  
Best Local Similarity 60.0%; Pred. No. 8.77e+00;  
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;  
  
Db 36 LPEESVLRSL 45  
QY 1 LPENNVLSPL 10  
  
RESULT 15  
ID YIJP\_ECOLI STANDARD; PRT; 577 AA.  
AC P32678;  
DT 01-OCT-1993 (Rel. 27, Created)  
DT 01-OCT-1993 (Rel. 27, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE HYPOTHETICAL 66.6 KD PROTEIN IN FRWD-PPC INTERGENIC REGION.  
GN YIJP.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Escherichia.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12 / MG1655;  
RX MEDLINE; 94089392.  
RA BLATTNER F.R., BURLAND V.D., PLUNKETT G. III, SOFIA H.J.,  
RA DANIELS D.L.;  
RT "Analysis of the Escherichia coli genome. IV. DNA sequence of the  
RT region from 89.2 to 92.8 minutes."  
RL Nucleic Acids Res. 21:5408-5417(1993).  
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE  
CC (PROBABLE).  
CC -!- SIMILARITY: BELONGS TO THE YHBY/YHJW/YIJP/YJDB FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; U00006; AAC43061.1; -  
DR EMBL; AE000469; AAC76937.1; -.

DR ECGENE; EGI1914; YIJP.  
KW Hypothetical protein; Transmembrane; Inner membrane.  
FT TRANSMEM 17 37 POTENTIAL.  
FT TRANSMEM 44 64 POTENTIAL.  
FT TRANSMEM 69 89 POTENTIAL.  
FT TRANSMEM 119 139 POTENTIAL.  
FT TRANSMEM 154 174 POTENTIAL.  
SQ SEQUENCE 577 AA; 66609 MW; 293C63FA CRC32;  
  
Query Match 70.1%; Score 47; DB 1; Length 577;  
Best Local Similarity 70.0%; Pred. No. 8.77e+00;  
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
  
Db 216 LNENNALPPL 225  
QY 1 LPENNVLSPL 10  
  
Search completed: Sat Apr 15 00:14:57 2000  
Job time : 41 secs.

THIS PAGE BLANK (USPTO)



\*\*\*\*\*  
M P S R L H  
\*\*\*\*\*  
(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Sat Apr 15 00:13:27 2000; MasPar time 3.35 Seconds  
119.769 Million cell updates/sec

Tabular output not generated.

Title: >US-08-452-843-13  
Description: (1-10) from US08452843.pep  
Perfect Score: 67  
Sequence: 1 LPENNVLSP 10

Scoring table: PAM 150  
Gap 15

Searched: 122810 seqs, 40068593 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: p1r2  
p1r1 2:p1r2 3:p1r3 4:p1r4

Statistics: Mean 23.293; Variance 28.545; scale 0.816

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES			
Result No.	Score	Query Match	Pred. No.
1	67	100.0	9.78e-04
2	67	100.0	9.78e-04
3	51	76.1	3.91e-00
4	51	76.1	3.91e-00
5	51	76.1	3.91e-00
6	51	76.1	3.91e-00
7	50	74.6	6.28e-00
8	49	73.1	1.00e+01
9	48	71.6	1.58e-01
10	48	71.6	1.58e-01
11	48	71.6	1.58e-01
12	48	71.6	1.58e-01
13	48	71.6	1.58e-01
14	47	70.1	2.49e-01
15	47	70.1	2.49e-01
16	47	70.1	2.49e-01
17	47	70.1	2.49e-01
18	47	70.1	2.49e-01
19	47	70.1	2.49e-01
20	46	68.7	3.89e-01
21	46	68.7	3.89e-01
22	46	68.7	3.89e-01
23	46	68.7	3.89e-01

24	46	68.7	386	2	S51648	cellular tumor antigen	3.89e-01
25	46	68.7	482	2	A34924	complement C3b/C4b re	3.89e-01
26	46	68.7	537	2	B46535	interleukin 2 recepto	6.03e-01
27	45	67.2	160	2	S44736	b0523.2 protein - Cae	6.03e-01
28	45	67.2	348	2	S44628	f22b7.1 protein - Cae	6.03e-01
29	45	67.2	605	2	H71303	hypothetical protein	6.03e-01
30	45	67.2	640	1	QYCHGM	phosphoenolpyruvate c	6.03e-01
31	45	67.2	860	2	S55543	reverse transcriptase	6.03e-01
32	45	67.2	908	2	S07649	gene coi intron 1 pro	6.03e-01
33	45	67.2	1124	1	GNLJFP	pol polyprotein - fel	6.03e-01
34	45	67.2	1124	2	S23820	pol polyprotein - fel	6.03e-01
35	45	67.2	1252	2	S77037	hypothetical protein	6.03e-01
36	45	67.2	3716	2	E70969	probable PPE protein	6.03e-01
37	44	65.7	255	2	A62965	conserved hypotheticala	9.27e-01
38	44	65.7	262	2	S60213	fomC protein - Strept	9.27e-01
39	44	65.7	273	2	D71276	conserved hypotheticala	9.27e-01
40	44	65.7	304	1	CRVZ7P	cell surface-binding	9.27e-01
41	44	65.7	452	2	S77538	serine proteinase (EC	9.27e-01
42	44	65.7	795	1	QRRTG	glucocorticoid recept	9.27e-01
43	44	65.7	806	2	G54109	dimethylsulfoxide red	9.27e-01
44	44	65.7	867	2	H70411	alanyl-tRNA synthetas	9.27e-01
45	44	65.7	2946	2	T00867	cell division control	9.27e-01

ALIGNMENTS

RESULT 1

ENTRY DNHU53 #type complete

TITLE cellular tumor antigen p53 - human

ALTERNATE\_NAMES cellular phosphoprotein p53; oncoprotein p53; transformation suppressor p53; tumor suppressor p53

ORGANISM #formal\_name Homo sapiens #common\_name man

DATE 05-Oct-1988 #sequence\_revision 18-Nov-1994 #text\_change 26-Feb-1999

ACCESSIONS A25224; A43073; J04336; S40773; S42669; A22837; A55060; A25397; B25397; S42452; S42453; I38082; I38083; I38084; I38085; I38086; I38087; I38088; I38089; I38090; I38091; I38092; I38093; A44905; I58354; I78850; I52681; S60153

REFERENCE A25224

#authors Lamb, P.; Crawford, L.

#journal Mol. Cell. Biol. (1986) 6:1379-1385

#title Characterization of the human p53 gene.

#cross-references M01D:87064416

#accession A25224

#molecule\_type DNA

#residues 1-393 #label LAM

#cross-references EMBL:X01405; GB:M13121; GB:N00032; NID:g189460; PID:g386994

REFERENCE J04336

#authors Buchman, V.L.; Chumakov, P.M.; Ninkina, N.N.; Samarina, O.P.; Georgiev, G.P.

#journal Gene (1988) 70:245-252

#title A variation in the structure of the protein-coding region of the human p53 gene.

#cross-references M01D:89108008

#accession A43073

#molecule\_type DNA

#residues 1-393 #label BUC1

#cross-references EMBL:M22898; NID:g189474

#note this 72-Arg allele appears to be about 5 times more frequent than the 72-Pro allele

REFERENCE J04336

#molecule\_type DNA

#residues 1-71,'P',73-393 #label BUC2

#cross-references EMBL:M22898; NID:g189474; PID:g189476

#note this 72-Pro allele was found in both normal and malignant cell lines

REFERENCE S40773

#authors Chumakov, P.M.; Almazov, V.P.; Jenkins, J.R.

#submission submitted to the EMBL Data Library, August 1990

#accession S40773

#molecule\_type DNA

#residues 1-393 #label CHU

```

#cross-references EMBL:X54156; NID:g35213; PID:g35214
REFERENCE
#accession S42669
#authors Matlashewski, G.; Lamb, P.; Pim, D.; Peacock, J.; Crawford, L.; Benchimol, S.
#journal EMBO J. (1984) 3:3257-3262
#title Isolation and characterization of a human p53 cDNA clone: expression of the human p53 gene.
#cross-references MUID:85126934
#accession S42669
#molecule_type mRNA
#residues 101-393 ##label MKI1
##cross-references EMBL:X01405; NID:g35215; PID:g642241
REFERENCE
#accession A22837
#authors Zakut-Houri, R.; Blenz-Tadmor, B.; Givol, D.; Oren, M.
#journal EMBO J. (1985) 4:1251-1255
#title Human p53 cellular tumor antigen: cDNA sequence and expression in COS cells.
#cross-references MUID:85230577
#accession A22837
#molecule_type mRNA
#residues 1-71,'P',73-393 ##label ZAK
##cross-references EMBL:X02469; EMBL:M60950; NID:g35209; PID:g35210
REFERENCE
#accession A5060
#authors Hariow, E.; Williamson, N.M.; Ralston, R.; Helfman, D.M.; Adams, T.E.
#journal Mol. Cell. Biol. (1985) 5:1601-1610
#title Molecular cloning and in vitro expression of a cDNA clone for human cellular tumor antigen p53.
#cross-references MUID:85267676
#accession A5060
#molecule_type mRNA
#residues 1-71,'P',73-272,'H',274-393 ##label HAR
##cross-references GB:K03199; NID:g189478; PID:g189479
##experimental_source clone pR4-2, cell line A431
REFERENCE
#accession A3086
#authors Harris, N.; Brill, E.; Shohat, O.; Prokocimer, M.; Wolf, D.; Arai, N.; Rotter, V.
#journal Mol. Cell. Biol. (1986) 6:4650-4656
#title Molecular basis for heterogeneity of the human p53 protein.
#cross-references MUID:87069826
#accession A25397
#molecule_type mRNA
#residues 1-78,'T',80-393 ##label HAR1
##cross-references EMBL:M14694; NID:g339813; PID:g339814
##experimental_source clone p53-H-1, transformed hybridoma SV-80 cell line
#accession B25397
#molecule_type mRNA
#residues 1-71,'P',73-78,'T',80-393 ##label HAR2
##cross-references EMBL:M14695; NID:g339815; PID:g339816
##experimental_source clone p53-H-19, transformed hybridoma SV-80 cell line
REFERENCE
#accession S42452
#authors Matlashewski, G.J.; Tuck, S.; Pim, D.; Lamb, P.; Schneider, J.; Crawford, L.V.
#journal Mol. Cell. Biol. (1987) 7:961-963
#title Primary structure polymorphism at amino acid residue 72 of human p53.
#cross-references MUID:87144273
#accession S42452
#molecule_type mRNA; DNA
#residues 66-71,'P',73-79 ##label MKI2
##experimental_source clone lambda C113
##note 72-Cys was also found, and appears to represent a polymorphism
#accession S42453
#molecule_type mRNA; DNA
#residues 66-79 ##label MKI3
##experimental_source clone J6K
REFERENCE
#accession I38082
#authors Farrell, P.J.; Allan, G.J.; Shanahan, F.; Vousden, K.H.; Crook, T.
#journal EMBO J. (1991) 10:2879-2887

```

```
#title      p53 is frequently mutated in Burkitt's lymphoma cell lines.
#cross-references OMIM:92007731
#accession  I38082
##molecule_type mRNA
##residues translated from GB/EMBL/DBJ
1-189 'LLSILSEKVCISVSNWTEFLDIVWCMPSRLRLALT',
'VPSSTTTCVTPVNAH' ##label F01
##cross-references EMBL:X60010; NID:g506432; PID:g506433
##note      deletion of a C nucleotide causes a frameshift at
              position 566
#accession  I38083
##status    translated from GB/EMBL/DBJ
##molecule_type mRNA
##residues 1-192, 'R', 194-393 ##label F02
##cross-references EMBL:X60011; NID:g506434; PID:g506435
#accession  I38084
##status    translated from GB/EMBL/DBJ
##molecule_type mRNA
##residues 1-393 ##label F03
##cross-references EMBL:X60012; NID:g506436; PID:g506437
#accession  I38085
##status    translated from GB/EMBL/DBJ
##molecule_type mRNA
##residues 1-245, 'T', 247-393 ##label F04
##cross-references EMBL:X60013; NID:g506438; PID:g506439
#accession  I38086
##status    translated from GB/EMBL/DBJ
##molecule_type mRNA
##residues 1-236, 'I', 238-393 ##label F05
##cross-references EMBL:X60014; NID:g506440; PID:g506441
#accession  I38087
##status    translated from GB/EMBL/DBJ
##molecule_type mRNA
##residues 1-247, 'Q', 249-393 ##label F06
##cross-references EMBL:X60015; NID:g506442; PID:g506443
#accession  I38088
##status    translated from GB/EMBL/DBJ
##molecule_type mRNA
##residues 1-71, 'P', 73-237, 'Y', 239-393 ##label F07
##cross-references EMBL:X60016; NID:g506444; PID:g506445
#accession  I38089
##status    translated from GB/EMBL/DBJ
##molecule_type mRNA
##residues 1-247, 'Q', 249-393 ##label F08
##cross-references EMBL:X60017; NID:g506446; PID:g506447
#accession  I38090
##status    translated from GB/EMBL/DBJ
##molecule_type mRNA
##residues 1-71, 'P', 73-162, 'H', 164-393 ##label F09
##cross-references EMBL:X60018; NID:g506448; PID:g506449
#accession  I38091
##status    translated from GB/EMBL/DBJ
##molecule_type mRNA
##residues 1-212, 'Q', 214-393 ##label F10
##cross-references EMBL:X60019; NID:g506450; PID:g506451
#accession  I38092
##status    translated from GB/EMBL/DBJ
##molecule_type mRNA
##residues 1-253, 'D', 255-393 ##label F11
##cross-references EMBL:X60020; NID:g506452; PID:g506453
##note      all sequences submitted to the EMBL/GenBank/DBJ
```

REFERENCE  
I38093  
#authors Jureal, P.A.; Barrett, J.C.; Wiseman, R.W.  
#journal Nucleic Acids Res. (1991) 19:6977  
#title An Alu polymorphism intragenic to the TP53 gene.  
#cross-references MIM:9210726  
#accession I38093  
#status translated from GB/EMBL/DBJ  
#molecule\_type DNA  
#residues 1-393  
#label FUT  
#cross-references EMBL:X54156; NID:g35213; PID:g35214  
REFERENCE  
A41905

```

#authors      Yamada, Y.; Yoshida, T.; Hayashi, K.; Sekiya, T.; Yokota, J.;
               Hirohashi, S.; Nakatani, K.; Nakano, H.; Sugimura, T.;
               Terada, M.
#journal      Cancer Res. (1991) 51:5800-5805
#title        p53 gene mutations in gastric cancer metastases and in
               gastric cancer cell lines derived from metastases.
#cross-references MUID:92034678
#accession    A44905
...
Note: remainder of annotations omitted.

Query Match      100.0%; Score 67; DB 1; Length 393;
Best Local Similarity 100.0%; Pred. No. 9.78e-04;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 26 LPENNVSPL 35
|||||
QY 1 LPENNVSPL 10

RESULT 2
ENTRY   S06594 #type complete
TITLE   cellular tumor antigen p53 - green monkey
ORGANISM #formal_name Cercopithecus aethiops #common_name green
          monkey, grivet
DATE    28-Feb-1990 #sequence_revision 28-Feb-1990 #text_change
          08-Sep-1997
ACCESSIONS S06594
REFERENCE   S06594
#authors   Rigaudy, P.; Eckhart, W.
#journal   Nucleic Acids Res. (1989) 17:8375
#title     Nucleotide sequence of a cDNA encoding the monkey cellular
           phosphoprotein p53.
#cross-references MUID:90045967
#accession S06594
           ##molecule_type mRNA
           ##residues 1-393 ##label RIG
           ##cross-references EMBL:X16384; NID:g22795; PID:g22796
CLASSIFICATION #superfamily cellular tumor antigen p53
KEYWORDS        apoptosis; cell division control; DNA binding; homotetramer;
               nucleus; phosphoprotein; transcription regulation; tumor
               suppressor; zinc

FEATURE
176,179,238,242 #binding_site zinc (Cys, His, Cys, Cys) #status
                predicted\
392 #binding_site phosphoryl-RNA (Ser) (covalent) #status
                predicted\
SUMMARY #length 393 #molecular-weight 43696 #checksum 4263

Query Match      100.0%; Score 67; DB 2; Length 393;
Best Local Similarity 100.0%; Pred. No. 9.78e-04;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 26 LPENNVSPL 35
|||||
QY 1 LPENNVSPL 10

RESULT 3
ENTRY   D70955 #type complete
TITLE   hypothetical protein Rv3603c - Mycobacterium tuberculosis
          (strain H37RV)
ORGANISM #formal_name Mycobacterium tuberculosis
DATE    17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change
          17-Jul-1998
ACCESSIONS D70955
REFERENCE   A70500
#authors   Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher,
           C.; Harris, D.; Gordon, S.V.; Eiglmeyer, K.; Gas, S.; Barry
           III, C.E.; Tekala, F.; Badcock, K.; Basham, D.; Brown, D.;
           Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.;
           Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;
           Hornsby, T.; Jagels, K.; Krogh, A.; McLean, J.; Moule, S.;

```

```

Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.;
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.;
Skellton, S.; Squares, S.; Squires, R.; Sulston, J.E.;
Taylor, K.; Whitehead, S.; Barrell, B.G.
Nature (1998) 393:537-544
Deciphering the biology of Mycobacterium tuberculosis from
the complete genome sequence.
#cross-references MUID:98295987
#accession D70955
#status preliminary; nucleic acid sequence not shown;
translation not shown
           ##molecule_type DNA
           ##residues 1-303 ##label COL
           ##cross-references GB:Z95557; GB:AL123456; NID:G3242276; PID:G316966;
           PID:G2113973
           ##experimental_source strain H37Rv
GENETICS
#gene Rv3603C
SUMMARY #length 303 #molecular-weight 31104 #checksum 129

Query Match      76.1%; Score 51; DB 2; Length 303;
Best Local Similarity 60.0%; Pred. No. 3.91e+00;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 55 LPDTPVLPPL 64
|||
QY 1 LPENNVSPL 10

RESULT 4
ENTRY   JC6176 #type complete
TITLE   tumor suppressor protein p53 - Chinese hamster
ORGANISM #formal_name Crictetus griseus #common_name Chinese hamster
DATE    11-Apr-1997 #sequence_revision 09-May-1997 #text_change
          08-Sep-1997
ACCESSIONS JC6176
REFERENCE   JC6176
#authors   Lee, H.; Lerner, J.M.; Hamlin, J.L.
#journal   Gene (1997) 184:177-183
#title     Cloning and characterization of Chinese hamster p53 cDNA.
#cross-references MUID:97183659
#contents   liver
#accession JC6176
           ##molecule_type mRNA
           ##residues 1-393 ##label LEE
           ##cross-references GB:U50395; NID:G1842229; PID:G1842230
COMMENT    This protein is a multimer, it plays the central role in a complex
           DNA damage-sensing network. It binds to replication factor and
           TATA-binding protein, and affects DNA replication, transcription,
           and recombination by protein/protein interactions.

GENETICS
#gene p53
CLASSIFICATION #superfamily cellular tumor antigen p53
KEYWORDS        liver; tumor
SUMMARY #length 393 #molecular-weight 43362 #checksum 4043

Query Match      76.1%; Score 51; DB 2; Length 393;
Best Local Similarity 80.0%; Pred. No. 3.91e+00;
Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 26 LPPNNVSTL 35
|||||
QY 1 LPENNVSPL 10

RESULT 5
ENTRY   JH0633 #type complete
TITLE   cellular tumor antigen p53 - golden hamster
ALTERNATE_NAMES tumor-suppressor protein p53
ORGANISM #formal_name Mesocricetus auratus #common_name golden hamster
DATE    17-Aug-1992 #sequence_revision 17-Aug-1992 #text_change
          08-Sep-1997
ACCESSIONS JH0633

```

```

REFERENCE
#authors      JH0633
#journal      Legros, Y.; McIntyre, P.; Soussi, T.
#title        Gene (1992) 112:247-250
#             The cDNA cloning and immunological characterization of
#             hamster p53.
#cross-references MUID:92210007
#accession      JH0633
##molecule_type mRNA
##residues      1-396 #label LEG
##cross-references GB:M75144; NID:g191414; PID:g191415
##experimental_source kidney, strain MP1

GENETICS
#gene          p53
#classification superfamily cellular tumor antigen p53
#keywords       apoptosis; cell division control; DNA binding; homotrimer;
#               nucleus; phosphoprotein; transcription regulation; tumor
#               suppressor; zinc
#feature        179,182,241,245 #binding_site zinc (Cys, His, Cys, Cys) #status
#               predicted\
#               predicted
#summary         length 396 #molecular-weight 43631 #checksum 6617
#               76.1%; Score 51; DB 2; Length 396;
#               Best Local Similarity 80.0%; Pred. No. 3.91e+00;
#               Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 26 LPNNVLSL 35
|| |||||
QY 1 LPNNVLSPL 10

RESULT 6
ENTRY S26849 #type complete
TITLE DNA-directed RNA polymerase (EC 2.7.7.6) II largest chain -
        fission yeast (Schizosaccharomyces pombe)
ORGANISM #formal_name Schizosaccharomyces pombe
DATE 25-Feb-1994 #sequence_revision 10-Nov-1995 #text_change
        12-Sep-1997
ACCESSIONS S26849
REFERENCE S26849
#authors      Azuma, Y.; Yamagishi, M.; Ueshima, R.; Ishihama, A.
#journal      Nucleic Acids Res. (1991) 19:461-468
#title        Cloning and sequence determination of the Schizosaccharomyces
#               pombe rpb1 gene encoding the largest subunit of RNA
#               polymerase II
#cross-references MUID:91187661
#accession     S26849
##molecule_type DNA
##residues      1-1752 #label AZU
##cross-references ENBL:X56564; NID:g5054; PID:g5055
##note          the authors did not translate the codon for residue 1464

GENETICS
#introns        5/3; 14/2; 38/3; 64/1; 84/1; 119/3
#classification #superfamily human DNA-directed RNA polymerase II largest
#               chain
#keywords        DNA binding; nucleotidyltransferase; tandem repeat;
#               transcription; zinc finger
#summary         length 1752 #molecular-weight 194161 #checksum 8039
#               76.1%; Score 51; DB 2; Length 1752;
#               Best Local Similarity 70.0%; Pred. No. 3.91e+00;
#               Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 202 LPEKRLSPL 211
|||: ||||
QY 1 LPNNVLSPL 10

RESULT 7
ENTRY S03785 #type complete
TITLE dimethylsulfoxide reductase (EC 1.8.-.-) chain A precursor,

```

```

anaerobic - Escherichia coli
#formal_name Escherichia coli
#journal      28-Feb-1990 #sequence_revision 28-Feb-1990 #text_change
#             05-Dec-1998
ACCESSIONS S03785; E64828
REFERENCE S03784
#authors      Bilous, P.T.; Cole, S.T.; Anderson, W.F.; Weiner, J.H.
#journal      Mol. Microbiol. (1988) 2:785-795
#title        Nucleotide sequence of the dmsABC operon encoding the
#               anaerobic dimethylsulphoxide reductase of Escherichia coli.
#cross-references MUID:89096500
#accession     S03785
##status        not compared with conceptual translation
##molecule_type DNA
##residues      1-785 #label BIL
##cross-references EMBL:J03412; NID:g145754; PID:g145755
##experimental_source strain C600
##note          part of this sequence, including the amino end of the
#               mature protein, was confirmed by protein sequencing
REFERENCE A64720
#authors      Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.;
#             Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.;
#             Rode, C.K.; Mayhew, G.E.; Gregor, J.; Davis, N.W.; Shao,
#             Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao,
#             Y.
#journal      Science (1997) 277:1453-1462
#title        The complete genome sequence of Escherichia coli K-12.
#cross-references MUID:97426617
#accession     E64828
##status        nucleic acid sequence not shown; translation not shown
##molecule_type DNA
##residues      1-785 #label BLAT
##cross-references GB:AE000191; GB:U00096; NID:g1787115; PID:g1787121;
#               UWGP:b0894
##experimental_source strain K-12, substrain MG1655

GENETICS
#gene          dmsA
#map_position   20 min
#function        heterotrimer; chains A, B, and C
#description     terminal reductase during anaerobic growth on various
#               sulfoxide and N-oxide compounds
#note            chain A binds molybdopterin, chain B is an electron transfer
#               protein and chain C an integral membrane protein
#keywords        4Fe-4S; heterotrimer; iron-sulfur protein; metalloprotein;
#               molybdenum; oxidoreductase
#feature         1-16 #domain signal sequence #status predicted #label SIG\
#               17-785 #product anaerobic dimethylsulfoxide reductase chain A
#               #status experimental #label MAT\
#               34,38,42,75 #binding_site iron-sulfur clusters (Cys) (covalent)
#               #status predicted
#summary         length 785 #molecular-weight 87448 #checksum 5497
#               74.6%; Score 50; DB 2; Length 785;
#               Best Local Similarity 60.0%; Pred. No. 6.28e+00;
#               Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 637 LPEGDVLDPL 646
|||: |||
QY 1 LPNNVLSPL 10

RESULT 8
ENTRY S77690 #type complete
TITLE probable membrane protein YOL075c - yeast (Saccharomyces
#               cerevisiae)
ALTERNATE_NAMES hypothetical protein Oll25; hypothetical protein Oll130;
#               hypothetical protein YOL074c
ORGANISM #formal_name Saccharomyces cerevisiae
DATE 21-Apr-1997 #sequence_revision 09-May-1997 #text_change
#               10-Jul-1998
ACCESSIONS S77690; S66767; S66768

```

```
REFERENCE S66756
#authors Alexandrakaki, D.; Katsoulou, C.; Tzermia, M.
#submission submitted to the Protein Sequence Database, July 1996
#accession S77690
#molecule_type DNA
##residues 1-1294 ##label ALE
##cross-references EMBL:274816; MIPS:YOL075c
##note this is a revision to the sequence from reference S66756

REFERENCE S66756
#authors Alexandrakaki, D.; Katsoulou, C.; Tzermia, M.
#submission submitted to the Protein Sequence Database, July 1996
#accession S66767
#molecule_type DNA
##residues 1-179, 'TTRTGVFLVVKRED' ##label ALW
##cross-references EMBL:274816
##experimental_source strain S288C
##note this sequence has been revised in reference S77690
##note this was assumed to be protein YOL074c

#accession S66768
#molecule_type DNA
##residues 200-1294 ##label ALF
##cross-references EMBL:274817
##experimental_source strain S288C
##note this sequence has been revised in reference S77690
##note this was assumed to be the complete sequence of protein YOL075c

GENETICS
#map_position 15L
#note YOL075c
CLASSIFICATION #superfamily unassigned ATP-binding cassette proteins;
ATP-binding cassette homology
P-loop; transmembrane protein
KEYWORDS
FEATURE
45-263 #domain ATP-binding cassette homology #label ABC1\
62-69 #region nucleotide-binding motif A (P-loop)\
376-392 #domain transmembrane #status predicted #label TM1\
469-485 #domain transmembrane #status predicted #label TM2\
496-512 #domain transmembrane #status predicted #label TM3\
606-622 #domain transmembrane #status predicted #label TM4\
710-916 #domain ATP-binding cassette homology #label ABC2\
727-734 #region nucleotide-binding motif A (P-loop)\
1042-1058 #domain transmembrane #status predicted #label TM5\
1125-1141 #domain transmembrane #status predicted #label TM6\
1177-1193 #domain transmembrane #status predicted #label TM7\
1269-1285 #domain transmembrane #status predicted #label TM8
#length 1294 #molecular-weight 145156 #checksum 3044

Query Match 73.1%; Score 49; DB 2; Length 1294;
Best Local Similarity 66.7%; Pred. No. 1.00e+01;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 126 LPOODVLSP 134
|||:||||
QY 1 LPENNVLSP 9

RESULT 9
ENTRY
TITLE
ALTERNATE_NAMES #type complete
ORGANISM hypothetical protein YDR179c - yeast (Saccharomyces cerevisiae)
#formal_name Saccharomyces cerevisiae
DATE 13-Jan-1995 #sequence_revision 10-Feb-1995 #text_change
12-Dec-1997

ACCESSIONS S49775
REFERENCE
#authors Murphy, L.; Harris, D.E.
#submission submitted to the EMBL Data Library, November 1994
#accession S49775
#molecule_type DNA
##residues 1-162 #label MUR
##cross-references EMBL:246727; NID:g1289283; PID:g223643; PID:g1289294; MIPS:YDR179c
```

```
GENETICS
#map_position 4R
SUMMARY #length 162 #molecular-weight 19476 #checksum 688

Query Match 71.6%; Score 48; DB 2; Length 162;
Best Local Similarity 60.0%; Pred. No. 1.58e+01;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 46 LPENILTSL 55
||||:|:|
QY 1 LPENNVLSP 10

RESULT 10
ENTRY
TITLE #type complete
ALTERNATE_NAMES tau-protein kinase (EC 2.7.1.135) homolog - common tobacco
glycogen synthase kinase 3 homolog; protein kinase GSK-3
ORGANISM #homolog; protein kinase shaggy homolog
#formal_name Nicotiana tabacum common_name common tobacco
DATE 14-Jul-1995 #sequence_revision 21-Jul-1995 #text_change
08-Sep-1997
ACCESSIONS S52095; S42085
REFERENCE S52095
#authors Einzenberger, E.; Eller, N.; Heberle-Bors, E.; Vicente, O.
#journal Biochim. Biophys. Acta (1995) 1260:315-319
#title Isolation and expression during pollen development of a tobacco cDNA clone encoding a protein kinase homologous to shaggy/glycogen synthase kinase-3.
#cross-references MIM:95178552
#accession S52095
#status preliminary
##molecule_type mRNA
##residues 1-409 ##label EIN
##cross-references EMBL:X77763; NID:g456355; PID:g456356
CLASSIFICATION #superfamily kinase-related transforming protein; protein
kinase homology
KEYWORDS ATP; phosphotransferase; serine/threonine-specific protein
kinase
FEATURE
71-332 #domain protein kinase homology #label KIM\
79-87 #region protein kinase ATP-binding motif\
102 #active_site Lys #status predicted
SUMMARY #length 409 #molecular-weight 46308 #checksum 7036

Query Match 71.6%; Score 48; DB 2; Length 409;
Best Local Similarity 60.0%; Pred. No. 1.58e+01;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 367 LPNGRVLPLP 376
|||:||||
QY 1 LPENNVLSP 10

RESULT 11
ENTRY
TITLE #type complete
ORGANISM probable esterase - Mycobacterium tuberculosis (strain H37RV)
#formal_name Mycobacterium tuberculosis
DATE 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change
17-Jul-1998

ACCESSIONS D70712
REFERENCE
#authors Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.V.; Eigmeier, K.; Basham, D.; Brown, D.; III, C.E.; Tekala, F.; Badcock, K.; Basham, D.; Brown, D.; Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Hornsby, T.; Jagels, K.; Krogh, A.; McLean, J.; Moule, S.; Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Squires, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
#journal Nature (1998) 393:537-544
#title Deciphering the biology of Mycobacterium tuberculosis from
```

```

the complete genome sequence.
#cross-references MUID:98295987
#accession D70712
##status preliminary; nucleic acid sequence not shown;
translation not shown
##molecule_type DNA
##residues 1-429 ##label COL
##cross-references GB:Z79701; GB:AL123456; NID:g3261635; PID:g264126;
PID:g1524244
##experimental_source strain H37Rv
GENETICS
#gene lipL
SUMMARY
#length 429 #molecular-weight 45815 #checksum 6197
Query Match 71.6%; Score 48; DB 2; Length 429;
Best Local Similarity 60.0%; Pred. No. 1.58e+01;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
Db 326 LPDNNLVPL 335
QY 1 LPENNVSPL 10

RESULT 12
ENTRY F69378 #type complete
TITLE conserved hypothetical protein AF1030 - Archaeoglobus
fulgidus
ORGANISM #formal_name Archaeoglobus fulgidus
DATE 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change
13-Sep-1998
F69378
A69250
#authors Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson,
K.E.; Ketchum, K.A.; Dodson, R.J.; Gwinn, M.; Hickey, E.K.;
Peterson, J.D.; Richardson, D.L.; Kerlavage, A.R.; Graham,
D.E.; Kyrpides, N.C.; Fleischmann, R.D.; Quackenbush, J.;
Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.;
Dougherty, B.A.; McKenny, K.; Adams, M.D.; Loftus, B.;
Peterson, S.; Reich, C.I.; McNeil, L.K.; Radger, J.H.;
Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman,
J.F.; McDonald, L.; Utterback, T.; Cotton, M.D.; Spriggs,
T.; Artlich, P.; Kaine, B.P.; Sykes, S.M.; Sadov, P.W.;
D'Andrea, K.P.; Bowman, C.; Fujii, C.; Garland, S.A.;
Mason, T.M.; Olsen, G.J.; Fraser, C.M.; Smith, H.O.; Woese,
C.R.; Venter, J.C.
#journal Nature (1997) 390:364-370
#title The complete genome sequence of the hyperthermophilic,
sulfate-reducing archaeon Archaeoglobus fulgidus.
#cross-references MUID:98049343
#accession F69378
##status preliminary; nucleic acid sequence not shown;
translation not shown
##molecule_type DNA
##residues 1-527 ##label KLE
##cross-references GB:AE001032; GB:AE000782; NID:g2689355; PID:g2649564;
TIGR:AF1030
CLASSIFICATION #superfamily conserved hypothetical protein MJ1429
SUMMARY #length 527 #molecular-weight 58876 #checksum 1184
Query Match 71.6%; Score 48; DB 2; Length 527;
Best Local Similarity 50.0%; Pred. No. 1.58e+01;
Matches 5; Conservative 5; Mismatches 0; Indels 0; Gaps 0;
Db 113 LADDVLSL 122
QY 1 LPENNVSPL 10

RESULT 13
ENTRY S75175 #type complete
TITLE hypothetical protein sl11945 - Synechocystis sp. (strain PCC
6803)
ORGANISM #formal_name Synechocystis sp.

the complete genome sequence.
#cross-references MUID:98295987
#accession D70712
##status preliminary; nucleic acid sequence not shown;
translation not shown
##molecule_type DNA
##residues 1-429 ##label COL
##cross-references GB:Z79701; GB:AL123456; NID:g3261635; PID:g264126;
PID:g1524244
##experimental_source strain H37Rv
GENETICS
#gene lipL
SUMMARY
#length 429 #molecular-weight 45815 #checksum 6197
Query Match 71.6%; Score 48; DB 2; Length 429;
Best Local Similarity 60.0%; Pred. No. 1.58e+01;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
Db 326 LPDNNLVPL 335
QY 1 LPENNVSPL 10

RESULT 12
ENTRY F69378 #type complete
TITLE conserved hypothetical protein AF1030 - Archaeoglobus
fulgidus
ORGANISM #formal_name Archaeoglobus fulgidus
DATE 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change
13-Sep-1998
F69378
A69250
#authors Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson,
K.E.; Ketchum, K.A.; Dodson, R.J.; Gwinn, M.; Hickey, E.K.;
Peterson, J.D.; Richardson, D.L.; Kerlavage, A.R.; Graham,
D.E.; Kyrpides, N.C.; Fleischmann, R.D.; Quackenbush, J.;
Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.;
Dougherty, B.A.; McKenny, K.; Adams, M.D.; Loftus, B.;
Peterson, S.; Reich, C.I.; McNeil, L.K.; Radger, J.H.;
Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman,
J.F.; McDonald, L.; Utterback, T.; Cotton, M.D.; Spriggs,
T.; Artlich, P.; Kaine, B.P.; Sykes, S.M.; Sadov, P.W.;
D'Andrea, K.P.; Bowman, C.; Fujii, C.; Garland, S.A.;
Mason, T.M.; Olsen, G.J.; Fraser, C.M.; Smith, H.O.; Woese,
C.R.; Venter, J.C.
#journal Nature (1997) 390:364-370
#title The complete genome sequence of the hyperthermophilic,
sulfate-reducing archaeon Archaeoglobus fulgidus.
#cross-references MUID:98049343
#accession F69378
##status preliminary; nucleic acid sequence not shown;
translation not shown
##molecule_type DNA
##residues 1-527 ##label KLE
##cross-references GB:AE001032; GB:AE000782; NID:g2689355; PID:g2649564;
TIGR:AF1030
CLASSIFICATION #superfamily conserved hypothetical protein MJ1429
SUMMARY #length 527 #molecular-weight 58876 #checksum 1184
Query Match 71.6%; Score 48; DB 2; Length 527;
Best Local Similarity 50.0%; Pred. No. 1.58e+01;
Matches 5; Conservative 5; Mismatches 0; Indels 0; Gaps 0;
Db 113 LADDVLSL 122
QY 1 LPENNVSPL 10

RESULT 13
ENTRY S75175 #type complete
TITLE hypothetical protein sl11945 - Synechocystis sp. (strain PCC
6803)
ORGANISM #formal_name Synechocystis sp.

PCC 6803
25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change
04-Sep-1998
#accessions S75175
#reference S74322
#authors Kaneo, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.;
Nakamura, Y.; Miyajima, N.; Hirose, M.; Sugita, M.;
Sasanoto, S.; Kimura, T.; Hosouchi, T.; Matsuno, A.;
Muraki, A.; Nakazaki, N.; Naruo, K.; Okumura, S.; Shimpo,
S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.;
Yasuda, M.; Tabata, S.
#journal DNA Res. (1996) 3:109-136
#title Sequence analysis of the genome of the unicellular
cyanobacterium Synechocystis sp. PCC6803. II. Sequence
determination of the entire genome and assignment of
potential protein-coding regions.
#cross-references MUID:97061201
#accession S75175
##status preliminary
##molecule_type DNA
##residues 1-640 ##label KAN
##cross-references EMBL:D90903; GB:AB001339; NID:g1652127; PID:d1017822;
PID:g1652165
##note the nucleotide sequence was submitted to the EMBL Data
Library, June 1996
CLASSIFICATION #superfamily hypothetical protein C2814
SUMMARY #length 640 #molecular-weight 69327 #checksum 6213
Query Match 71.6%; Score 48; DB 2; Length 640;
Best Local Similarity 70.0%; Pred. No. 1.58e+01;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
Db 580 LMDNNLVPL 589
QY 1 LPENNVSPL 10

RESULT 14
ENTRY S45299 #type complete
TITLE hypothetical protein - Trypanosoma brucei
ORGANISM #formal_name Trypanosoma brucei
DATE 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change
08-Jan-1995
#accessions S45299
#reference S45299
#authors Glauser, A.; Braun, R.
#journal Biochim. Biophys. Acta (1994) 1218:99-101
#title TUBIS, a fossilized retroposon in the tubulin gene cluster of
Trypanosoma brucei.
#cross-references MUID:94250702
#accession S45299
##status preliminary
##molecule_type DNA
##residues 1-101 ##label GLA
##cross-references EMBL:X62164
SUMMARY #length 101 #molecular-weight 11327 #checksum 995
Query Match 70.1%; Score 47; DB 2; Length 101;
Best Local Similarity 50.0%; Pred. No. 2.49e+01;
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;
Db 9 VP0DVLGSL 18
QY 1 LPENNVSPL 10

RESULT 15
ENTRY JC4042 #type complete
TITLE lexA protein - Aeromonas hydrophila
ORGANISM #formal_name Aeromonas hydrophila
DATE 13-Jun-1995 #sequence_revision 14-Jul-1995 #text_change
08-Sep-1997
#accessions JC4042

```

```

REFERENCE      JC4042
#authors      Riera, J.; Barbe, J.
#journal      Gene (1995) 154:71-75
#title        Cloning, sequence and regulation of expression of the lexA
               gene of Aeromonas hydrophila.
#cross-references MUID:95172406
#accession    JC4042
               #molecule_type DNA
               ##residues 1-207 #label RIE
               ##cross-references EMBL:X77263; NID:g840713; PID:g840714
GENETICS
#gene         lexA
CLASSIFICATION #superfamily lexA repressor
KEYWORDS       SOS response; transcription regulation
SUMMARY        #length 207 #molecular-weight 22898 #checksum 4250

Query Match    70.1%; Score 47; DB 2; Length 207;
Best Local Similarity 70.0%; Pred.No. 2.49e+01;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 173 LPNEELSPI 182
    ||||: |||:
Qy 1 LPENNVLSPL 10

```

Search completed: Sat Apr 15 00:13:59 2000  
 Job time : 32 secs.

THIS PAGE BLANK (USTIU)



\*\*\*\*\*  
[W][I][L][D][E][R][N][E][S][S]  
\*\*\*\*\*  
(TM)  
\*\*\*\*\*

Release 3.1A John F. Collins, BioComputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Sat Apr 15 00:15:15 2000; Maspar time 7.27 Seconds  
Tabular output not generated. 95.367 Million cell updates/sec

Title: >US-08-452-843-13  
Description: (1-10) from US08452843.pep  
Sequence: 1 LPENNVSPL 10

Scoring table: PAM 150  
Gap 15

Searched: 225878 seqs, 69334122 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: sptrembl12  
1:sp\_archaea 2:sp\_bacteria 3:sp\_fungi 4:sp\_human  
5:sp\_invertebrate 6:sp\_mammal 7:sp\_mhc 8:sp\_organelle  
9:sp\_phase 10:sp\_plant 11:sp\_rodent 12:sp\_unclassified  
13:sp\_vertebrate 14:sp\_virus

Statistics: Mean 23.257; Variance 23.542; scale 0.988  
Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES					Pred. No.	
Result No.	Score	Query %	Length	Description	ID	
1	67	100.0	393	P53 TRANSFORMATION SUP	Q15087	5.94e-05
2	67	100.0	393	P53 TRANSFORMATION SUP	Q15088	5.94e-05
3	67	100.0	393	P53 TRANSFORMATION SUP	Q15086	5.94e-05
4	67	100.0	393	CELLULAR TUMOR ANTIGEN	Q16810	5.94e-05
5	67	100.0	393	CELLULAR TUMOR ANTIGEN	Q16807	5.94e-05
6	67	100.0	393	CELLULAR TUMOR ANTIGEN	Q16808	5.94e-05
7	67	100.0	393	P53 TRANSFORMATION SUP	Q16535	5.94e-05
8	67	100.0	393	CELLULAR TUMOR ANTIGEN	Q16809	5.94e-05
9	67	100.0	393	CELLULAR TUMOR ANTIGEN	Q16848	5.94e-05
10	67	100.0	393	CELLULAR TUMOR ANTIGEN	Q16811	5.94e-05
11	60	99.6	391	CELLULAR TUMOR ANTIGEN	Q36006	5.60e-03
12	53	79.1	729	BRCA1 ASSOCIATED PROTE	Q23560	3.98e-01
13	51	76.1	161	PT15.	Q04681	1.27e-00
14	51	76.1	303	HYPOTHETICAL 31.1 KD P	Q06279	1.27e-00
15	51	76.1	391	CELLULAR TUMOR ANTIGEN	Q3WUR6	1.27e-00
16	50	74.6	1911	COMPLEMENT RECEPTOR 1	Q29528	2.23e-00
17	49	73.1	199	HYPOTHETICAL 21.7 KD P	Q08233	3.90e-00
18	49	73.1	685	SYNAPTPODIN.	Q45271	3.90e-00
19	49	73.1	692	SYNAPTPODIN.	Q32327	3.90e-00
20	49	73.1	1494	SIMILAR TO THE HUMAN M	Q20943	3.90e-00

21	49	73.1	1525	5	Q94137	MULTIDRUG RESISTANCE R	3.90e+00
22	49	73.1	1540	5	Q94136	MULTIDRUG RESISTANCE R	3.90e+00
23	48	71.6	162	3	Q03981	HYPOTHETICAL 19.5 KD P	6.76e+00
24	48	71.6	409	10	Q40518	SHAGGY LIKE PROTEIN KI	6.76e+00
25	48	71.6	429	2	P71178	HYPOTHETICAL 45.8 KD P	6.76e+00
26	48	71.6	527	1	Q29232	CONSERVED HYPOTHETICAL	6.76e+00
27	48	71.6	960	5	Q9XXW0	ENDONUCLEASE AND REVER	6.76e+00
28	48	71.6	960	5	Q93137	REVERSE TRANSCRIPTASE.	6.76e+00
29	48	71.6	1072	4	Q9Y4G7	KIAA0319 PROTEIN.	6.76e+00
30	47	70.1	127	14	Q66936	POL POLYPROTEIN (FRAGM	1.16e+01
31	47	70.1	155	14	Q96774	POL POLYPROTEIN (FRAGM	1.16e+01
32	47	70.1	159	14	Q96776	POL POLYPROTEIN (FRAGM	1.16e+01
33	47	70.1	159	14	Q96772	POL POLYPROTEIN (FRAGM	1.16e+01
34	47	70.1	159	14	Q96761	POL POLYPROTEIN (FRAGM	1.16e+01
35	47	70.1	159	14	Q96767	POL POLYPROTEIN (FRAGM	1.16e+01
36	47	70.1	159	14	Q96766	POL POLYPROTEIN (FRAGM	1.16e+01
37	47	70.1	159	14	Q96764	POL POLYPROTEIN (FRAGM	1.16e+01
38	47	70.1	159	14	Q96765	POL POLYPROTEIN (FRAGM	1.16e+01
39	47	70.1	159	14	Q96750	POL POLYPROTEIN (FRAGM	1.16e+01
40	47	70.1	159	14	Q96779	POL POLYPROTEIN (FRAGM	1.16e+01
41	47	70.1	1086	14	Q84809	POL POLYPROTEIN.	1.16e+01
42	47	70.1	1123	14	Q66933	POLYMERASE (FRAGMENT).	1.16e+01
43	47	70.1	1150	14	P90246	POL POLYPROTEIN.	1.16e+01
44	47	70.1	1568	4	Q95785	HUMAN HOMOLOG OF MUS M	1.16e+01
45	47	70.1	2021	2	Q52657	190-KDA ANTIGEN (ROMPA	1.16e+01

ALIGNMENTS

RESULT 1		PRELIMINARY;		PRT;		393 AA.	
ID	Q15087						
AC	Q15087						
DT	01-NOV-1996	(TEMBLrel. 01, Created)					
DT	01-NOV-1996	(TEMBLrel. 01, Last sequence update)					
DT	01-NOV-1999	(TEMBLrel. 12, Last annotation update)					
DE	P53 TRANSFORMATION SUPPRESSOR (FRAGMENT).						
GN	P53.						
OS	Homo sapiens (Human).						
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;						
OC	Eutheria; Primates; Catarrhini; Hominidae; Homo.						
RN	[1]						
RP	SEQUENCE FROM N.A.						
RX	MEDLINE; 92007731.						
RA	FARRELL P.J., ALLAN G., SHANAHAN F., VOUSDEN K.H., CROOK T.;						
RT	"p53 is frequently mutated in Burkitt's lymphoma cell lines."						
RL	EMBO J. 10:2875-2887(1991).						
DR	ENBL; X60014; CAA42629.1; -.						
DR	HSSP; P04637; 1SAH.						
DR	PFAM; PF00870; P53; 1.						
FT	VARIANT 237 237						
FT	NON-TER 393 393						
SQ	SEQUENCE 393 AA; 43694 MW; 9BB81992 CRC32;						

Query Match		100.0%;		Score 67; DB 4; Length 393;	
Best Local Similarity		100.0%;		Pred. No. 5.94e-05;	
Matches		10; Conservative		0; Mismatches 0; Indels 0; Gaps 0;	
Db	26 LPENNVSPL 35				
QY	1 LPENNVSPL 10				
RESULT 2					
ID	Q15088	PRELIMINARY;		PRT;	393 AA.
AC	Q15088:				
DT	01-NOV-1996	(TEMBLrel. 01, Created)			
DT	01-NOV-1996	(TEMBLrel. 01, Last sequence update)			
DT	01-NOV-1999	(TEMBLrel. 12, Last annotation update)			
DE	P53 TRANSFORMATION SUPPRESSOR (FRAGMENT).				
GN	P53.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;				
OC	Eutheria; Primates; Catarrhini; Hominidae; Homo.				

**-!- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT**

DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
 DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)  
 DE CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).  
 GN P53.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 92007731.  
 RA FARRELL P.J., ALLAN G., SHANAHAN F., VOUSDEN K.H., CROOK T.;  
 RL EMBO J. 10:2879-2887(1991).  
 CC -1- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT  
 CC PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL  
 CC CYCLE REGULATION, IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY  
 CC REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED  
 CC FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF  
 CC CYCLIN-DEPENDENT KINASES (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 CC EMBL: X60018; CAA42633.1; -.  
 DR HSSP: P04637; 1SAH.  
 DR PROSITE: PS00348; P53; 1.  
 DR PFAM: PF00870; P53; 1.  
 DR KW Anti-oncogene; DNA-binding; Transcription regulation; Activator;  
 KW Nuclear protein; Phosphorylation.  
 FT VARIANT 163 163 H -> Y.  
 FT NON\_TER 393 393  
 SQ SEQUENCE 393 AA; 43627 MW; AFD8A9E3 CRC32;

Query Match 100.0%; Score 67; DB 4; Length 393;  
 Best Local Similarity 100.0%; Pred. No. 5.94e-05;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 26 LPENNVLSP 35  
 QY 1 LPENNVLSP 10

RESULT 7  
 ID Q16535 PRELIMINARY; PRT; 393 AA.  
 AC Q16535;  
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
 DT 01-NOV-1999 (TrEMBLrel. 01, Last sequence update)  
 DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)  
 DE P53 TRANSFORMATION SUPPRESSOR (FRAGMENT).  
 GN P53.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 92007731.  
 RA FARRELL P.J., ALLAN G., SHANAHAN F., VOUSDEN K.H., CROOK T.;  
 RL EMBO J. 10:2879-2887(1991).  
 CC -1- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT  
 CC PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL  
 CC CYCLE REGULATION, IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY  
 CC REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED  
 CC FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF  
 CC CYCLIN-DEPENDENT KINASES (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 CC EMBL: X60017; CAA42632.1; -.  
 DR HSSP: P04637; 1SAH.  
 DR PROSITE: PS00348; P53; 1.  
 DR PFAM: PF00870; P53; 1.  
 DR KW Anti-oncogene; DNA-binding; Transcription regulation; Activator;  
 KW Nuclear protein; Phosphorylation.  
 FT VARIANT 248 248 Q -> R.  
 FT NON\_TER 393 393  
 SQ SEQUENCE 393 AA; 43684 MW; 239818A9 CRC32;

Query Match 100.0%; Score 67; DB 4; Length 393;  
 Best Local Similarity 100.0%; Pred. No. 5.94e-05;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 26 LPENNVLSP 35  
 QY 1 LPENNVLSP 10

RESULT 8  
 ID Q16809 PRELIMINARY; PRT; 393 AA.  
 AC Q16809;  
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
 DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)  
 DE CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).  
 GN P53.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 92007731.  
 RA FARRELL P.J., ALLAN G., SHANAHAN F., VOUSDEN K.H., CROOK T.;  
 RL EMBO J. 10:2879-2887(1991).  
 CC -1- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT  
 CC PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL  
 CC CYCLE REGULATION, IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY  
 CC REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED  
 CC FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF  
 CC CYCLIN-DEPENDENT KINASES (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 CC EMBL: X60019; CAA42634.1; -.  
 DR HSSP: P04637; 1SAH.  
 DR PROSITE: PS00348; P53; 1.  
 DR PFAM: PF00870; P53; 1.  
 DR KW Anti-oncogene; DNA-binding; Transcription regulation; Activator;  
 KW Nuclear protein; Phosphorylation.  
 FT VARIANT 213 213 Q -> R.  
 FT NON\_TER 393 393  
 SQ SEQUENCE 393 AA; 43684 MW; CB70BD7F CRC32;

Query Match 100.0%; Score 67; DB 4; Length 393;  
 Best Local Similarity 100.0%; Pred. No. 5.94e-05;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 26 LPENNVLSP 35  
 QY 1 LPENNVLSP 10

RESULT 9  
 ID Q16848 PRELIMINARY; PRT; 393 AA.  
 AC Q16848;  
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
 DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)  
 DE CELLULAR TUMOR ANTIGEN P53.  
 GN TP53.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 87089826.  
 RA HARRIS N., BRILL E., SHOHAT O., PROKOCIMER M., WOLF D., ARAI N.,  
 RA ROTTER V.;  
 RT "Molecular basis for heterogeneity of the human p53 protein.";  
 RL Mol. Cell. Biol. 6:4650-4656(1986).  
 CC -1- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT  
 CC PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL  
 CC CYCLE REGULATION, IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY  
 CC REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED  
 CC FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF  
 CC CYCLIN-DEPENDENT KINASES (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 CC EMBL: M14694; AAA61211.1; -.  
 DR HSSP: P04637; 1TSR.  
 DR PROSITE: PS00348; P53; 1.  
 DR PFAM: PF00870; P53; 1.  
 DR PRINTS: PR00386; P53SUPPRESSR.

KW Nuclear protein; Phosphorylation; Anti-oncogene; DNA-binding;  
 KW Transcription regulation; Activator.  
 SQ SEQUENCE 393 AA; 43723 MW; DA7D302F CRC32;

Query Match 100.0%; Score 67; DB 4; Length 393;  
 Best Local Similarity 100.0%; Pred. No. 5.94e-05;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 26 LPENNVSPL 35  
 QY 1 LPENNVSPL 10

RESULT 10  
 ID Q16811 PRELIMINARY; PRT; 393 AA.  
 AC Q16811;  
 DT 01-NOV-1996 (TRENBLrel. 01, Created)  
 DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)  
 DT 01-NOV-1999 (TRENBLrel. 12, Last annotation update)  
 DE CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 85126934.  
 RA MATLASHESKI G., LAMB P., PIM D., PEACOCK J., CRAWFORD L.,  
 RA BENCHIMOL S.;  
 RT "Isolation and characterization of a human p53 cDNA clone: expression  
 of the human p53 gene.";  
 RL EMBO J. 3:3257-3262(1984).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 87064416.  
 RA LAMB P., CRAWFORD L.;  
 RT "Characterization of the human p53 gene.";  
 RL Mol. Cell. Biol. 6:1379-1385(1986).  
 CC -!- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT  
 PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL  
 CYCLE REGULATION, IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY  
 REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED  
 FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF  
 CYCLIN-DEPENDENT KINASES (BY SIMILARITY).  
 CC -!- SUBCELLULAR LOCATION: NUCLEAR.

DR EMBL; M13121; AAA59987.1; JOINED.  
 DR EMBL; M13112; AAA59987.1; JOINED.  
 DR EMBL; M13113; AAA59987.1; JOINED.  
 DR EMBL; M13114; AAA59987.1; JOINED.  
 DR EMBL; M13115; AAA59987.1; JOINED.  
 DR EMBL; M13116; AAA59987.1; JOINED.  
 DR EMBL; M13117; AAA59987.1; JOINED.  
 DR EMBL; M13118; AAA59987.1; JOINED.  
 DR EMBL; M13119; AAA59987.1; JOINED.  
 DR EMBL; M13120; AAA59987.1; JOINED.  
 DR HSSP; P04637; ITSR.  
 DR PROSITE; PS00348; P53; 1.  
 DR PFAM; PF00870; P53; 1.  
 KW Repeat: Tumor antigen; Anti-oncogene; DNA-binding;  
 KW Transcription regulation; Activator; Nuclear protein; Phosphorylation.  
 FT NON\_TER 393 393  
 SQ SEQUENCE 393 AA; 43698 MW; 3EA71431 CRC32;

Query Match 100.0%; Score 67; DB 4; Length 393;  
 Best Local Similarity 100.0%; Pred. No. 5.94e-05;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 26 LPENNVSPL 35  
 QY 1 LPENNVSPL 10

RESULT 11  
 ID O36006 PRELIMINARY; PRT; 391 AA.

O36006;  
 DT 01-JAN-1998 (TRENBLrel. 05, Created)  
 DT 01-JAN-1998 (TRENBLrel. 05, Last sequence update)  
 DT 01-NOV-1999 (TRENBLrel. 12, Last annotation update)  
 DE CELLULAR TUMOR ANTIGEN P53.  
 GN P53.  
 OS Marmota monax (Woodchuck).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Rodentia; Sciurognathi; Scuriidae; Scuriinae; Marmota.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 97376996.  
 RA FEITELSON M.A., RANGANATHAN P.N., CLAYTON M.M., ZHANG S.M.;  
 RT "Partial characterization of the woodchuck tumor suppressor, p53, and  
 its interaction with woodchuck hepatitis virus X antigen in  
 hepatocarcinogenesis.";  
 RL Oncogene 15:327-336(1997).  
 CC -!- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT  
 PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL  
 CYCLE REGULATION, IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY  
 REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED  
 FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF  
 CYCLIN-DEPENDENT KINASES (BY SIMILARITY).  
 CC -!- SUBCELLULAR LOCATION: NUCLEAR.  
 DR EMBL; AJ001022; CAA04478.1; -.  
 DR HSSP; P04637; ITSR.  
 DR PROSITE; PS00348; P53; 1.  
 DR PFAM; PF00870; P53; 1.  
 DR PRINTS; PR00386; P53SUPPRESSR.  
 KW Anti-oncogene; DNA-binding; transcription regulation; Activator;  
 KW Nuclear protein; Phosphorylation.  
 SQ SEQUENCE 391 AA; 43468 MW; 95FAB8F2 CRC32;

Query Match 89.6%; Score 60; DB 6; Length 391;  
 Best Local Similarity 90.0%; Pred. No. 5.60e-03;  
 Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 26 LPENNVSPL 35  
 QY 1 LPENNVSPL 10

RESULT 12  
 ID Q92560 PRELIMINARY; PRT; 729 AA.

AC Q92560;  
 DT 01-FEB-1997 (TRENBLrel. 02, Created)  
 DT 01-NOV-1998 (TRENBLrel. 08, Last sequence update)  
 DT 01-NOV-1999 (TRENBLrel. 12, Last annotation update)  
 DE BRCA1 ASSOCIATED PROTEIN 1 (MYELOBLAST KIAA0272).  
 GN. BAP1 OR KIAA0272.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 98187701.  
 RA JENSEN D.E., PROCTOR M., MARQUIS S.T., GARDNER H.P., HA S.,  
 RA CHODOSH L.A., ISHOV A.M., TOMMERUP N., VISSING H., SEKIDO Y.,  
 RA MINNA J., BORODOVSKY A., SCHULTZ D.C., WILKINSON K.D., MAUL G.G.,  
 RA BARLEV N., BERGER S., PRENDERGAST G.C., RAUSCHER F.J. III.;  
 RT "BAP1: a novel ubiquitin hydrolase which binds to the BRCA1 RING  
 finger and enhances BRCA1-mediated cell growth suppression.";  
 RL Oncogene 16:1097-1112(1998).  
 RN [2]  
 RP SEQUENCE OF 4-729 FROM N.A.  
 RC TISSUE-BRAIN;  
 RX MEDLINE; 97191544.  
 RA NAGASE T., SEKI N., ISHIKAWA K., OHIRA M., KAWARABAYASI Y., OHARA O.,  
 RA TANAKA A., KOTANI H., MIYAJIMA N., NOMURA N.;  
 RT "Prediction of the coding sequences of unidentified human genes. VI.  
 The coding sequences of 80 new genes (K1AA0201-K1AA0280) deduced by  
 analysis of cDNA clones from cell line KG-1 and brain.";  
 RL DNA Res. 3:321-329(1996).

```
DR EMBL; AF045581; AAC15970.1; -.
DR EMBL; D87462; BAA13401.1; -.
DR HSSP; P15374; UCH.
DR PAM; PF01088; UCH; 1.
SQ SEQUENCE 729 AA; 80361 MW; 038968F7 CRC32;

Query Match 79.1%; Score 53; DB 4; Length 729;
Best Local Similarity 70.0%; Pred. No. 3.98e+01;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 564 LAEDGVLSPL 573
Qy 1 LPENNVLSP 10

RESULT 13
ID 004681 PRELIMINARY; PRT; 161 AA.
AC 004681;
DT 01-JUL-1997 (TRENBLrel. 04, Created)
DT 01-JUL-1997 (TRENBLrel. 04, Last sequence update)
DT 01-NOV-1999 (TRENBLrel. 12, Last annotation update)
DE PT15.
OS Lycopersicon esculentum (Tomato).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC eumolliphytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Asteridae; euasterids I; Solanales; Solanaceae;
OC Solanum.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 97357308.
RA ZHOU J., TANG X., MARTIN G.B.;
RT "The Pto Kinase conferring resistance to tomato bacterial speck
RT disease interacts with proteins that bind a cis-element of
RT pathogenesis-related genes.";
RL EMBO J. 16:3207-3218(1997).
DR EMBL; U89256; AAC49740.1; -.
DR MENDEL; 16334; Lyces; 2475; 16334.
DR PFAM; PF00847; AP2-domain; 1.
DR PRINTS; PR00367; ETHRSPELENT.
SQ SEQUENCE 161 AA; 18051 MW; 9ADA5570 CRC32;

Query Match 76.1%; Score 51; DB 10; Length 161;
Best Local Similarity 70.0%; Pred. No. 1.27e+00;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 43 LPRNNILRPL 52
Qy 1 LPENNVLSP 10

RESULT 14
ID 006279 PRELIMINARY; PRT; 303 AA.
AC 006279;
DT 01-JUL-1997 (TRENBLrel. 04, Created)
DT 01-JUL-1997 (TRENBLrel. 04, Last sequence update)
DT 01-NOV-1998 (TRENBLrel. 08, Last annotation update)
DE HYPOTHETICAL 31.1 KD PROTEIN.
GN MTCY07H7B.19.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-H37RV;
RA DEVLIN K., CHURCHER C.M.;
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-H37RV;
RA BARRELL B.G., RAJANDREAM M.A.;
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
```

```
RC STRAIN-H37RV;
RX MEDLINE; 96181548.
RA PHILIPP W.J., POULET S., EIGLMEIER K., PASCOPELLA L.,
RA BALASUBRAMANIAN V., HEYM B., BERGH S., BLOOM B.R., JACOBS W.R. JR.,
RA COLE S.T.;
RT "An integrated map of the genome of the tubercle bacillus,
RT Mycobacterium tuberculosis H37RV, and comparison with Mycobacterium
RT leprae.";
RL EMBL; Z95557; CAB08941.1; -.
DR KW Hypothetical protein.
KW SEQUENCE 303 AA; 31104 MW; 4FAFFIEC CRC32;

Query Match 76.1%; Score 51; DB 2; Length 303;
Best Local Similarity 60.0%; Pred. No. 1.27e+00;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 55 LPDTPVLPL 64
Qy 1 LPENNVLSP 10

RESULT 15
ID 09WUR6 PRELIMINARY; PRT; 391 AA.
AC 09WUR6;
DT 01-NOV-1999 (TRENBLrel. 12, Created)
DT 01-NOV-1999 (TRENBLrel. 12, Last sequence update)
DT 01-NOV-1999 (TRENBLrel. 12, Last annotation update)
DE CELLULAR TUMOR ANTIGEN P53.
GN P53.
OS Cavia porcellus (Guinea pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Hystricognathi; Caviidae; Cavia.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-SPLEEN;
RX MEDLINE; 99265972.
RA D'ERCHIA A.M., PESOLE G., TULLIO A., SACCONI C., SBISA E.;
RT "Guinea pig p53 mRNA: identification of new elements in coding and
RT untranslated regions and their functional and evolutionary
RT implications.";
RL Genomics 58:50-64(1999).
CC -1- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT
CC PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL
CC CYCLE REGULATION, IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY
CC REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED
CC FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF
CC CYCLIN-DEPENDENT KINASES (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
DR EMBL; AJ009673; CAB43196.1; -.
DR PROSITE; PS00348; P53; 1.
KW Anti-oncogene; DNA-binding; Transcription regulation; Activator;
KW Nuclear protein; Phosphorylation.
KW SEQUENCE 391 AA; 43288 MW; BFD34AB4 CRC32;

Query Match 76.1%; Score 51; DB 11; Length 391;
Best Local Similarity 100.0%; Pred. No. 1.27e+00;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 26 LPENNVLSP 33
Qy 1 LPENNVLSP 8
```

Search completed: Sat Apr 15 00:16:47 2000  
Job time : 92 secs.

**THIS PAGE BLANK (USPTO)**

\*\*\*\*\*  
[M][P][S][R][C][H] (TM)  
\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Sat Apr 15 00:27:50 2000; Maspar time 3.20 Seconds  
81.488 Million cell updates/sec

Tabular output not generated.

Title: >US-08-452-843-15  
Description: (1-11) from US08452843.pap  
Perfect Score: 86  
Sequence: 1 SPALNMFQCOL 11

Scoring table: PAM 150  
Gap 15

Searched: 188963 seqs, 23686106 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: a-geneseq36  
1:geneseq36

Statistics: Mean 17.949; Variance 50.752; scale 0.354

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES				Pred. No.	
Result No.	Score	Query Match	Description	DB ID	
1	86	100.0	p53 protein residues 1	1	W03363
2	86	100.0	Human p53 protein vari	1	W28484
3	86	100.0	Human p53 protein vari	1	W28483
4	86	100.0	Human p53 protein vari	1	W28495
5	86	100.0	Human p53 protein vari	1	W28496
6	86	100.0	Human p53 protein vari	1	W28498
7	86	100.0	Human p53 protein vari	1	W28497
8	86	100.0	Chimeric p53 protein	1	W13962
9	86	100.0	Human p53 protein vari	1	W28493
10	86	100.0	Human p53 protein vari	1	W28494
11	86	100.0	Chimeric p53 protein	1	W13960
12	86	100.0	Chimeric p53 protein	1	W13961
13	86	100.0	Human p53 protein vari	1	W28479
14	86	100.0	Modified p53 variant (	1	W13954
15	86	100.0	Modified p53 variant p	1	W13972
16	86	100.0	Modified p53 variant p	1	W13975
17	86	100.0	Modified p53 variant p	1	W13976
18	86	100.0	Modified p53 variant p	1	W13971
19	86	100.0	Human p53 protein vari	1	W28480
20	86	100.0	Human p53 protein vari	1	W28482
21	86	100.0	Human p53 protein vari	1	W28481
22	86	100.0	Human p53 protein vari	1	W28490
23	86	100.0	Human p53 protein vari	1	W28489

24	86	100.0	393	1	Y03191	Amino acid sequence of	5.46e-03
25	86	100.0	393	1	W84270	Human p53 protein.	5.46e-03
26	86	100.0	393	1	W69218	Human p53 mutant 1.	5.46e-03
27	86	100.0	393	1	W69217	Human wild-type p53 pr	5.46e-03
28	86	100.0	393	1	W57244	Human p53 protein SEQ	5.46e-03
29	86	100.0	393	1	W05345	Human p53 mutant R273H	5.46e-03
30	86	100.0	393	1	W05347	Human p53 mutant R248Q	5.46e-03
31	86	100.0	393	1	W13968	Modified p53 variant p	5.46e-03
32	86	100.0	393	1	W13970	Modified p53 variant p	5.46e-03
33	86	100.0	393	1	W25155	Human p53 variant foun	5.46e-03
34	86	100.0	393	1	W05349	Human p53 mutant R273C	5.46e-03
35	86	100.0	393	1	W02617	Human p53 tumour suppr	5.46e-03
36	86	100.0	393	1	W05348	Human p53 mutant R282W	5.46e-03
37	86	100.0	393	1	W13978	Human tumour-derived p	5.46e-03
38	86	100.0	393	1	W13952	Human tumour-derived p	5.46e-03
39	86	100.0	393	1	W13951	Human tumour-derived p	5.46e-03
40	86	100.0	393	1	W13949	T284R modified human p	5.46e-03
41	86	100.0	401	1	W28488	Human p53 protein vari	5.46e-03
42	86	100.0	402	1	W13965	Chimeric p53 protein.	5.46e-03
43	86	100.0	406	1	W13966	Chimeric p53 protein.	5.46e-03
44	86	100.0	411	1	W13967	Chimeric p53 protein.	5.46e-03
45	86	100.0	535	1	W28491	Human p53 protein vari	5.46e-03

ALIGNMENTS

RESULT 1  
ID W03363 standard; peptide; 28 AA.  
AC W03363;  
DT 10-MAR-1997 (first entry)  
DE P53 protein residues 124-151.  
KW Cytotoxic T lymphocyte; CTL; epitope; p53; template;  
KW rathect library; pharmaceutical; vaccine; treatment; prevention;  
KW disease; malignancy; cancer.  
OS Homo sapiens.  
FH Key  
FT region  
FT 11..20  
FT /note= "cytotoxic T lymphocyte epitope"  
PN W09622067-A2.  
PD 25-JUL-1996.  
PF 15-DEC-1995; U16290.  
PR 27-DEC-1994; US-366332.  
PA (UNBI-) UNITED BIOMEDICAL INC.  
PI Kuebler PJ, Nixon DF;  
DR WPI; 96-354273/35.  
PT Rathect library of peptide(s) contg. an immuno:stimulatory CTL  
PT epitope - derived from longer template peptide, useful as  
PT pharmaceutical or vaccine against infectious disease or malignancy  
PS Claim 9; Pages 36-37; 60pp; English.  
CC The present peptide comprises residues 124-151 of the p53 protein,  
CC contains cytotoxic T lymphocyte (CTL) epitope and can be used as a  
CC template in the prepn. of a rathect library, comprising peptides  
CC contg. at least 1 immunostimulatory CTL epitope. Basically the  
CC distribution of amino acids at each position in the template is  
CC calculated, a rathect library constructed from the longer template  
CC peptide by sequentially rathecting it into the shorter rathect  
CC length and the peptides synthesised using standard solid phase  
CC methods. The library can be used in pharmaceuticals and vaccines  
CC for the treatment and/or prevention of diseases and malignancies  
CC associated with p53 mutation, e.g. cancer.  
CC Several epitopes can be incorporated into the same library, rather  
CC than using a mixt. of individually synthesised immunogenic  
CC peptides, which helps to overcome problems of genetic diversity  
CC and MHC restriction. The library may also include antigenic  
CC variations and escape mutations.  
SQ Sequence 28 AA;

Query Match 100.0%; Score 86; DB 1; Length 28;  
Best Local Similarity 100.0%; Pred. No. 5.46e-03;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Db 4 SPALNMFQCOL 14  
|||||||

```
QY 1 SPALNKMFCOL 11
RESULT 2
ID W28484 standard; Protein; 253 AA.
AC W28484;
DE 25-NOV-1997 (first entry)
DT 17-JUL-1996; F01111.
KW Human p53 protein variant V-367H.
KW Leucine zipper domain; LZD; oligomerisation domain; mutant; muten;
KW substitution; replacement; transactivation; viral protein VP16; HSV;
KW anti-oncogene; hyperproliferation; cancer; restenosis;
KW tumour suppression; apoptosis.
OS Chimeric - Homo sapiens.
OS Chimeric - Herpes simplex virus.
OS Synthetic.
FH Location/Qualifiers
FT misc_difference 189
FT /note= "Arg residue at position 182 of wild-type
FT p53 has been mutated to His"
FT
PN WO9704092-A1.
PD 06-FEB-1997.
PF 17-JUL-1996; F01111.
PR 19-JUL-1995; FR-008729.
PA (RHON ) RHONE POULENC RORER SA.
PI Bracco L, Conseiller E;
DR WPI; 97-132633/12.
PT New p53 variants e.g. with oligomerisation domain replaced by
PT leucine zipper - useful for treating hyper-proliferative disorders,
PT esp. cancer and restenosis
PS Claim 32; Page -; 133pp; French.
CC Claimed variants of protein p53 have at least part of the p53
CC transactivation domain (amino acids 1-74) deleted and replaced by
CC the transactivating domain (TD) from herpes simplex virus viral
CC protein VP16 (amino acids 411-490). The present sequence is that of
CC a specifically claimed p53 variant designated V-367 and comprising
CC the VP16 TD and amino acids 75-367 of human wild-type p53. The p53
CC variants are more active and more stable tumour suppressors and
CC apoptosis-inducing agents than wild-type p53 and are active where
CC the wild-type protein is not.
CC Sequence 253 AA;
QY 1 SPALNKMFCOL 11
Query Match 100.0%; Score 86; DB 1; Length 253;
Best Local Similarity 100.0%; Pred. No. 5.46e-03;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 135 SPALNKMFCOL 145
QY 1 SPALNKMFCOL 11
RESULT 3
ID W28483 standard; Protein; 253 AA.
AC W28483;
DE 25-NOV-1997 (first entry)
DT 17-JUL-1996; F01111.
KW Human p53 protein variant V-367 encoded by pC141.
KW Leucine zipper domain; LZD; oligomerisation domain; mutant; muten;
KW substitution; replacement; transactivation; viral protein VP16; HSV;
KW anti-oncogene; hyperproliferation; cancer; restenosis;
KW tumour suppression; apoptosis.
OS Chimeric - Homo sapiens.
OS Chimeric - Herpes simplex virus.
OS Synthetic.
FH Location/Qualifiers
FT misc_difference 189
FT /note= "Arg residue at position 182 of wild-type
FT p53 has been mutated to His"
FT
PN WO9704092-A1.
PD 06-FEB-1997.
PF 17-JUL-1996; F01111.
PR 19-JUL-1995; FR-008729.
PA (RHON ) RHONE POULENC RORER SA.
PI Bracco L, Conseiller E;
DR WPI; 97-132633/12.
PT New p53 variants e.g. with oligomerisation domain replaced by
PT leucine zipper - useful for treating hyper-proliferative disorders,
PT esp. cancer and restenosis
PS Claim 32; Page -; 133pp; French.
CC Claimed variants of protein p53 have at least part of the p53
CC transactivation domain (amino acids 1-74) deleted and replaced by
CC the transactivating domain (TD) from herpes simplex virus viral
CC protein VP16 (amino acids 411-490). The present sequence is that of
CC a specifically claimed p53 variant designated V-367 and comprising
CC the VP16 TD and amino acids 75-367 of human wild-type p53 (but with
CC Arg182 replaced by His). The p53 variants are more active and more
CC stable tumour suppressors and apoptosis-inducing agents than wild-type
CC p53 and are active where the wild-type protein is not.
CC (Note: this sequence does not appear in the specification and has
CC been produced by modifying the given sequence of variant V-367).
CC Sequence 253 AA;
QY 1 SPALNKMFCOL 11
Query Match 100.0%; Score 86; DB 1; Length 253;
Best Local Similarity 100.0%; Pred. No. 5.46e-03;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 135 SPALNKMFCOL 145
QY 1 SPALNKMFCOL 11
RESULT 4
ID W28495 standard; Protein; 319 AA.
AC W28495;
DE 25-NOV-1997 (first entry)
DT Human p53 protein variant 360-325 encoded by pC178.
KW Leucine zipper domain; LZD; oligomerisation domain; mutant; muten;
KW substitution; replacement; transactivation; viral protein VP16; HSV.
KW anti-oncogene; hyperproliferation; cancer; restenosis;
KW tumour suppression; apoptosis.
OS Homo sapiens.
OS Synthetic.
FH Location/Qualifiers
FT misc_difference 189
FT /note= "Arg residue at position 182 of wild-type
FT p53 has been mutated to His"
FT
PN WO9704092-A1.
PD 06-FEB-1997.
PF 17-JUL-1996; F01111.
PR 19-JUL-1995; FR-008729.
PA (RHON ) RHONE POULENC RORER SA.
PI Bracco L, Conseiller E;
DR WPI; 97-132633/12.
PT New p53 variants e.g. with oligomerisation domain replaced by
PT leucine zipper - useful for treating hyper-proliferative disorders,
PT esp. cancer and restenosis
PS Claim 38; Pages 92-94; 133pp; French.
CC Claimed variants of protein p53 have at least part of the
CC oligomerisation domain deleted and replaced by a leucine zipper
CC domain. The mutants preferably also have at least part of the p53
CC transactivation domain (amino acids 1-74) deleted and replaced by
CC the domain 325-360 of p53. The present sequence is that of a
CC specifically claimed p53 variant designated 360-325 and comprising
CC the 325-360 domain, amino acids 75-325 of human wild-type p53 and a
CC leucine zipper domain at the C-terminal. The p53 variants are
CC more active and more stable tumour suppressors and apoptosis-inducing
CC agents than wild-type p53 and are active where the wild-type protein
CC is not, i.e. they are not inactivated by dominant negative or oncogenic
CC mutants, nor by other cellular proteins (because the leucine zipper
CC domain prevents formation of inactive mixed oligomers).
CC Sequence 319 AA;
QY 1 SPALNKMFCOL 101
QY 1 SPALNKMFCOL 11
RESULT 5
ID W28496 standard; Protein; 319 AA.
AC W28496;
```



DT 25-NOV-1997 (first entry)  
DE Human p53 protein variant 360-325H.  
KW Leucine zipper domain; LZD; oligomerisation domain; mutant; muten;  
KW substitution; replacement; transactivation; viral protein VP16; HSV;  
KW anti-oncogene; hyperproliferation; cancer; restenosis;  
KW tumour suppression; apoptosis.  
OS Homo sapiens.  
OS Synthetic.  
FH Key  
FT Location/Qualifiers  
FT misc\_difference 145  
FT /note= "Arg residue at position 182 of wild-type  
FT p53 has been mutated to His"  
FT  
PD WO9704092-A1.  
PN 06-FEB-1997.  
PF 17-JUL-1996; F01111.  
PR 19-JUL-1995; FR-008729.  
PA (RHON ) RHONE POULENC RORER SA.  
PI Bracco L, Conseiller E;  
PI WPI; 97-132633/12.  
DR New p53 variants e.g. with oligomerisation domain replaced by  
PT leucine zipper - useful for treating hyper-proliferative disorders,  
PT esp. cancer and restenosis  
PS Claim 38; Page -: 133pp; French.  
CC Claimed variants of protein p53 have at least part of the  
CC oligomerisation domain deleted and replaced by a leucine zipper  
CC domain. The mutants preferably also have at least part of the p53  
CC transactivation domain (amino acids 1-74) deleted and replaced by  
CC the domain 325-360 of p53. The present sequence is that of a  
CC specifically claimed p53 variant designated 360-325H and comprising  
CC the 325-360 domain, separated from amino acids 75-325 of human  
CC wild-type p53 (but with Arg182 replaced by His) by a synthetic hinge  
CC sequence (Gly4Ser)3, and with a leucine zipper domain at the C-terminal.  
CC The p53 variants are more active and more stable tumour suppressors  
CC and apoptosis-inducing agents than wild-type p53 and are active where  
CC the wild-type protein is not, i.e. they are not inactivated by dominant  
CC negative or oncogenic mutants, nor by other cellular proteins (because  
CC the leucine zipper domain prevents formation of inactive mixed  
CC oligomers).  
CC (Note: this sequence does not appear in the specification and has  
CC been produced by modifying the given sequence of variant 360-325).  
SQ Sequence 319 AA;  
  
Query Match 100.0%; Score 86; DB 1; Length 319;  
Best Local Similarity 100.0%; Pred. No. 5.46e-03;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Db 91 SPALNMFQOL 101  
Qy 1 SPALNMFQOL 11  
|||||  
  
RESULT 6  
ID W28498 standard; Protein; 335 AA.  
AC W28498.  
DT 25-NOV-1997 (first entry)  
DE Human p53 protein variant 360h-325H.  
KW Leucine zipper domain; LZD; oligomerisation domain; mutant; muten;  
KW substitution; replacement; transactivation; hinge region;  
KW anti-oncogene; hyperproliferation; cancer; restenosis;  
KW tumour suppression; apoptosis.  
OS Homo sapiens.  
OS Synthetic.  
FH Key  
FT Location/Qualifiers  
FT region 39..53  
FT /label= hinge  
FT misc\_difference 161  
FT /note= "Arg residue at position 182 of wild-type  
FT p53 has been mutated to His"  
FT  
PD WO9704092-A1.  
PN 06-FEB-1997.  
PF 17-JUL-1996; F01111.  
PR 19-JUL-1995; FR-008729.  
PA (RHON ) RHONE POULENC RORER SA.  
PI Bracco L, Conseiller E;  
PI WPI; 97-132633/12.  
DR New p53 variants e.g. with oligomerisation domain replaced by  
PT leucine zipper - useful for treating hyper-proliferative disorders,  
PT esp. cancer and restenosis  
PS Claim 39; Pages 94-95; 133pp; French.  
CC Claimed variants of protein p53 have at least part of the  
CC oligomerisation domain deleted and replaced by a leucine zipper  
CC domain. The mutants preferably also have at least part of the p53  
CC transactivation domain (amino acids 1-74) deleted and replaced by  
CC the domain 325-360 of p53. The present sequence is that of a  
CC specifically claimed p53 variant designated 360h-325 and comprising  
CC the 325-360 domain, separated from amino acids 75-325 of human  
CC wild-type p53 by a synthetic hinge sequence (Gly4Ser)3, and with a  
CC leucine zipper domain at the C-terminal. The p53 variants are  
CC more active and more stable tumour suppressors and apoptosis-inducing  
CC agents than wild-type p53 and are active where the wild-type protein  
CC is not, i.e. they are not inactivated by dominant negative or oncogenic  
CC mutants, nor by other cellular proteins (because the leucine zipper  
CC domain prevents formation of inactive mixed oligomers).

DR WPI; 97-132633/12.  
PT New p53 variants e.g. with oligomerisation domain replaced by  
PT leucine zipper - useful for treating hyper-proliferative disorders,  
PT esp. cancer and restenosis  
PS Claim 39; Page -: 133pp; French.  
CC Claimed variants of protein p53 have at least part of the  
CC oligomerisation domain deleted and replaced by a leucine zipper  
CC domain. The mutants preferably also have at least part of the p53  
CC transactivation domain (amino acids 1-74) deleted and replaced by  
CC the domain 325-360 of p53. The present sequence is that of a  
CC specifically claimed p53 variant designated 360h-325H and comprising  
CC the 325-360 domain, separated from amino acids 75-325 of human  
CC wild-type p53 (but with Arg182 replaced by His) by a synthetic hinge  
CC sequence (Gly4Ser)3, and with a leucine zipper domain at the C-terminal.  
CC The p53 variants are more active and more stable tumour suppressors  
CC and apoptosis-inducing agents than wild-type p53 and are active where  
CC the wild-type protein is not, i.e. they are not inactivated by dominant  
CC negative or oncogenic mutants, nor by other cellular proteins (because  
CC the leucine zipper domain prevents formation of inactive mixed  
CC oligomers).  
CC (Note: this sequence does not appear in the specification and has  
CC been produced by modifying the given sequence of variant 360h-325).  
SQ Sequence 335 AA;  
  
Query Match 100.0%; Score 86; DB 1; Length 335;  
Best Local Similarity 100.0%; Pred. No. 5.46e-03;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Db 107 SPALNMFQOL 117  
Qy 1 SPALNMFQOL 11  
|||||  
  
RESULT 7  
ID W28497 standard; Protein; 335 AA.  
AC W28497.  
DT 25-NOV-1997 (first entry)  
DE Human p53 protein variant 360h-325 encoded by p53179.  
KW Leucine zipper domain; LZD; oligomerisation domain; mutant; muten;  
KW substitution; replacement; transactivation; hinge region;  
KW anti-oncogene; hyperproliferation; cancer; restenosis;  
KW tumour suppression; apoptosis.  
OS Homo sapiens.  
OS Synthetic.  
FH Key  
FT Location/Qualifiers  
FT region 39..53  
FT /label= hinge  
FT WO9704092-A1.  
PN 06-FEB-1997.  
PF 17-JUL-1996; F01111.  
PR 19-JUL-1995; FR-008729.  
PA (RHON ) RHONE POULENC RORER SA.  
PI Bracco L, Conseiller E;  
PI WPI; 97-132633/12.  
DR N-PSDB; T86224.  
DR New p53 variants e.g. with oligomerisation domain replaced by  
PT leucine zipper - useful for treating hyper-proliferative disorders,  
PT esp. cancer and restenosis  
PS Claim 39; Pages 94-95; 133pp; French.  
CC Claimed variants of protein p53 have at least part of the  
CC oligomerisation domain deleted and replaced by a leucine zipper  
CC domain. The mutants preferably also have at least part of the p53  
CC transactivation domain (amino acids 1-74) deleted and replaced by  
CC the domain 325-360 of p53. The present sequence is that of a  
CC specifically claimed p53 variant designated 360h-325 and comprising  
CC the 325-360 domain, separated from amino acids 75-325 of human  
CC wild-type p53 by a synthetic hinge sequence (Gly4Ser)3, and with a  
CC leucine zipper domain at the C-terminal. The p53 variants are  
CC more active and more stable tumour suppressors and apoptosis-inducing  
CC agents than wild-type p53 and are active where the wild-type protein  
CC is not, i.e. they are not inactivated by dominant negative or oncogenic  
CC mutants, nor by other cellular proteins (because the leucine zipper  
CC domain prevents formation of inactive mixed oligomers).

```
SQ Sequence 335 AA;
Query Match 100.0%; Score 86; DB 1; Length 335;
Best Local Similarity 100.0%; Pred. No. 5.46e-03;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 107 SPALNMFQOL 117
|||||
QY 1 SPALNMFQOL 11

RESULT 8
ID W13962 standard; Protein: 337 AA.
AC W13962;
DE 25-JUN-1997 (first entry)
KW Chimeric p53 protein.
KW p53; tumour suppressor; cancer; therapy; cell proliferation;
KW apoptosis; protein engineering; GCN4; DNA binding.
OS Chimeric Homo sapiens;
OS Chimeric synthetic.
FH Key Location/Qualifiers
FT region 1..300
FT /label= p53wt
FT /note= "amino acids 1-300 of wild-type p53"
FT region 301..305
FT /label= Linker
FT region 306..337
FT /label= GCN4
FT /note= "amino acids 250-281 of GCN4 LZ variant"
PN W09710843-A1.
PD 27-MAR-1997.
PR 20-SEP-1996; U15188.
PR 22-SEP-1995; US-004802.
PR 21-AUG-1996; US-697221.
PA (WIST-) WISTAR INST ANATOMY & BIOLOGY.
PI Halazonetis TD;
DR WPI; 97-202618/18.
PT R284K modified p53 protein having DNA binding ability - useful in
PT treatment of cancer
PS Disclosure: Refer to Page 8; 82pp; English.
CC Chimeric p53 constructs (W13956-57) comprise N-terminal portions
CC of human wild-type p53 tumour suppressor (see also W13948) linked
CC to a C-terminal portion of the LZ variant (see also W13955) of
CC GCN4 and, in some cases, the C-terminal portion of wild-type
CC p53. The chimeric proteins have DNA binding activity and can
CC replace lost or insufficient p53 function, providing the means for
CC pharmacological rescue of p53 function in cancer patients. Nucleic
CC acids coding for modified p53 constructs can be used for cancer
CC gene therapy.
SQ Sequence 337 AA;

Query Match 100.0%; Score 86; DB 1; Length 337;
Best Local Similarity 100.0%; Pred. No. 5.46e-03;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 127 SPALNMFQOL 137
|||||
QY 1 SPALNMFQOL 11

RESULT 9
ID W28493 standard; Protein: 353 AA.
AC W28493;
DE 25-NOV-1997 (first entry)
KW Human p53 protein variant 393-325 encoded by p53.
KW Leucine zipper domain; LZD; oligomerisation domain; mutant; mutagen;
KW anti-oncogene; hyperproliferation; cancer; restenosis;
KW tumour suppression; apoptosis.
OS Homo sapiens.
OS Synthetic.
PN W09704092-A1.
PD 06-FEB-1997.

Query Match 100.0%; Score 86; DB 1; Length 337;
Best Local Similarity 100.0%; Pred. No. 5.46e-03;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 127 SPALNMFQOL 137
|||||
QY 1 SPALNMFQOL 11

RESULT 9
ID W28493 standard; Protein: 353 AA.
AC W28493;
DE 25-NOV-1997 (first entry)
KW Human p53 protein variant 393-325 encoded by p53.
KW Leucine zipper domain; LZD; oligomerisation domain; mutant; mutagen;
KW anti-oncogene; hyperproliferation; cancer; restenosis;
KW tumour suppression; apoptosis.
OS Homo sapiens.
OS Synthetic.
PN W09704092-A1.
PD 06-FEB-1997.

Query Match 100.0%; Score 86; DB 1; Length 353;
Best Local Similarity 100.0%; Pred. No. 5.46e-03;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 125 SPALNMFQOL 135
|||||
QY 1 SPALNMFQOL 11

RESULT 10
ID W28494 standard; Protein: 353 AA.
AC W28494;
DE 25-NOV-1997 (first entry)
KW Human p53 protein variant 393-325H.
KW Leucine zipper domain; LZD; oligomerisation domain; mutant; mutagen;
KW substitution; replacement; transactivation; viral protein VP16; HSV;
KW anti-oncogene; hyperproliferation; cancer; restenosis;
KW tumour suppression; apoptosis.
OS Homo sapiens.
OS Synthetic.
FH Key Location/Qualifiers
FT misc_difference 179
FT /note= "Arg residue at position 182 of wild-type
FT p53 has been mutated to His"
PN W09704092-A1.
PD 06-FEB-1997.
PR 17-JUL-1996; F01111.
PR 19-JUL-1995; FR-008729.
PA (RHON) RHONE POULENC RORER SA.
PI Bracco L, Conseiller E;
DR WPI; 97-132633/12.
PT New p53 variants e.g. with oligomerisation domain replaced by
PT leucine zipper - useful for treating hyper-proliferative disorders,
PT esp. cancer and restenosis
PS Claim 37; Page 133pp; French.
CC Claimed variants of protein p53 have at least part of the
CC oligomerisation domain deleted and replaced by a leucine zipper
CC domain. The mutants preferably also have at least part of the p53
CC transactivation domain (amino acids 1-74) deleted and replaced by
CC the domain 325-393 of p53. The present sequence is that of
CC a specifically claimed p53 variant designated 393-325 and comprising
CC the 325-393 domain, amino acids 75-325 of human wild-type p53 and a
CC leucine zipper domain at the C-terminal. The p53 variants are
CC more active and more stable tumour suppressors and apoptosis-inducing
CC agents than wild-type p53 and are active where the wild-type protein
CC is not, i.e. they are not inactivated by dominant negative or oncogenic
CC mutants, nor by other cellular proteins (because the leucine zipper
CC domain prevents formation of inactive mixed oligomers).
SQ Sequence 353 AA;
```

```
CC oligomers).
CC (Note: this sequence does not appear in the specification and has
CC been produced by modifying the given sequence of variant 393-325).
SQ Sequence 353 AA;

Query Match 100.0%; Score 86; DB 1; Length 353;
Best Local Similarity 100.0%; Pred. No. 5.46e-03;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 125 SPALNMFQOL 135
QY 1 SPALNMFQOL 11

RESULT 11
ID W13960 standard; Protein; 359 AA.
AC W13960;
DT 25-JUN-1997 (first entry)
DE Chimeric p53 protein.
KW p53; tumour suppressor; cancer; therapy; cell proliferation;
KW apoptosis; protein engineering; GCN4; DNA binding.
OS Chimeric Homo sapiens;
OS Chimeric synthetic.
FH Key Location/Qualifiers
FT region 1..323
FT /label= p53wt
FT /note= "amino acids 1-323 of wild-type p53"
FT 324..326
FT /label= Linker
FT 327..359
FT /label= GCN4
FT /note= "amino acids 249-281 of GCN4 LZ variant"
PN WO9710843-A1.
PD 27-MAR-1997.
PE 20-SEP-1996; U15188.
PR 22-SEP-1995; US-004802.
PR 21-AUG-1996; US-697221.
PA (WIST-) WISTAR INST ANATOMY & BIOLOGY.
PI Halazonetis TD;
DR WPI; 97-202618/18.
PT R284K modified p53 protein having DNA binding ability - useful in
PT treatment of cancer
PS Disclosure; Refer to Page 8; 82pp; English.
CC Chimeric p53 constructs (W13956-67) comprise N-terminal portions
CC of human wild-type p53 tumour suppressor (see also W13948) linked
CC to a C-terminal portion of the LZ variant (see also W13955) of
CC GCN4 and, in some cases, the C-terminal portion of wild-type
CC p53. The chimeric proteins have DNA binding activity and can
CC replace lost or insufficient p53 function, providing the means for
CC pharmacological rescue of p53 function in cancer patients. Nucleic
CC acids coding for modified p53 constructs can be used for cancer
CC gene therapy.
CC Sequence 359 AA;

Query Match 100.0%; Score 86; DB 1; Length 359;
Best Local Similarity 100.0%; Pred. No. 5.46e-03;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 127 SPALNMFQOL 137
QY 1 SPALNMFQOL 11

RESULT 12
ID W13961 standard; Protein; 361 AA.
AC W13961;
DT 25-JUN-1997 (first entry)
DE Chimeric p53 protein.
KW p53; tumour suppressor; cancer; therapy; cell proliferation;
KW apoptosis; protein engineering; GCN4; DNA binding.
OS Chimeric Homo sapiens;
OS Chimeric synthetic.
FH Key Location/Qualifiers
FT region 1..323
FT /label= p53wt
FT /note= "amino acids 1-323 of wild-type p53"
FT 324..326
FT /label= Linker
FT 327..359
FT /label= GCN4
FT /note= "amino acids 249-281 of GCN4 LZ variant"
PN WO9710843-A1.
PD 27-MAR-1997.
PE 20-SEP-1996; U15188.
PR 22-SEP-1995; US-004802.
PR 21-AUG-1996; US-697221.
PA (WIST-) WISTAR INST ANATOMY & BIOLOGY.
PI Halazonetis TD;
DR WPI; 97-202618/18.
PT R284K modified p53 protein having DNA binding ability - useful in
PT treatment of cancer
PS Disclosure; Refer to Page 8; 82pp; English.
CC Chimeric p53 constructs (W13956-67) comprise N-terminal portions
CC of human wild-type p53 tumour suppressor (see also W13948) linked
CC to a C-terminal portion of the LZ variant (see also W13955) of
CC GCN4 and, in some cases, the C-terminal portion of wild-type
CC p53. The chimeric proteins have DNA binding activity and can
CC replace lost or insufficient p53 function, providing the means for
CC pharmacological rescue of p53 function in cancer patients. Nucleic
CC acids coding for modified p53 constructs can be used for cancer
CC gene therapy.
CC Sequence 359 AA;

Query Match 100.0%; Score 86; DB 1; Length 359;
Best Local Similarity 100.0%; Pred. No. 5.46e-03;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 127 SPALNMFQOL 137
QY 1 SPALNMFQOL 11

RESULT 13
ID W28479 standard; Protein; 363 AA.
AC W28479;
DT 25-NOV-1997 (first entry)
DE Human p53 protein variant V-325 encoded by pEC114.
KW Leucine zipper domain; LZD; oligomerisation domain; mutant; mutein;
KW substitution; replacement; transactivation; viral protein VP16; HSV;
KW anti-oncogene; hyperproliferation; cancer; restenosis;
KW tumour suppression; apoptosis.
OS Chimeric - Homo sapiens.
OS Chimeric - Homo sapiens.
OS Synthetic.
PN WO9704092-A1.
PD 06-FEB-1997.
PE 17-JUL-1996; F01111.
PR 19-JUL-1995; FR-008729.
PA (RHON ) RHONE POULENC RORER SA.
PI Bracco L, Conseiller E;
DR WPI; 97-132633/12.
DR N-PSDB; T86215.
PT New p53 variants e.g. with oligomerisation domain replaced by
PT leucine zipper - useful for treating hyper-proliferative disorders,
PT esp. cancer and restenosis
PS Claim 30; Pages 76-78; 133pp; French.
CC Claimed variants of protein p53 have at least part of the
CC oligomerisation domain deleted and replaced by a leucine zipper
CC domain. The mutants preferably also have at least part of the p53
CC transactivation domain (amino acids 1-74) deleted and replaced by
CC the transactivating domain (TD) from herpes simplex virus viral
CC protein VP16 (amino acids 411-490). The present sequence is that of
CC a specifically claimed p53 variant designated V-325 and comprising
CC the VP16 TD, amino acids 75-325 of human wild-type p53 and a
CC leucine zipper domain at the C-terminal. The p53 variants are
CC more active and more stable tumour suppressors and apoptosis-inducing
```

```
FT region 1..323
FT /label= p53wt
FT /note= "amino acids 1-323 of wild-type p53"
FT 324..329
FT /label= Linker
FT 330..361
FT /label= GCN4
FT /note= "amino acids 250-281 of GCN4 LZ variant"
PN WO9710843-A1.
PD 27-MAR-1997.
PE 20-SEP-1996; U15188.
PR 22-SEP-1995; US-004802.
PR 21-AUG-1996; US-697221.
PA (WIST-) WISTAR INST ANATOMY & BIOLOGY.
PI Halazonetis TD;
DR WPI; 97-202618/18.
PT R284K modified p53 protein having DNA binding ability - useful in
PT treatment of cancer
PS Disclosure; Refer to Page 8; 82pp; English.
CC Chimeric p53 constructs (W13956-67) comprise N-terminal portions
CC of human wild-type p53 tumour suppressor (see also W13948) linked
CC to a C-terminal portion of the LZ variant (see also W13955) of
CC GCN4 and, in some cases, the C-terminal portion of wild-type
CC p53. The chimeric proteins have DNA binding activity and can
CC replace lost or insufficient p53 function, providing the means for
CC pharmacological rescue of p53 function in cancer patients. Nucleic
CC acids coding for modified p53 constructs can be used for cancer
CC gene therapy.
CC Sequence 361 AA;

Query Match 100.0%; Score 86; DB 1; Length 361;
Best Local Similarity 100.0%; Pred. No. 5.46e-03;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 127 SPALNMFQOL 137
QY 1 SPALNMFQOL 11

RESULT 13
ID W28479 standard; Protein; 363 AA.
AC W28479;
DT 25-NOV-1997 (first entry)
DE Human p53 protein variant V-325 encoded by pEC114.
KW Leucine zipper domain; LZD; oligomerisation domain; mutant; mutein;
KW substitution; replacement; transactivation; viral protein VP16; HSV;
KW anti-oncogene; hyperproliferation; cancer; restenosis;
KW tumour suppression; apoptosis.
OS Chimeric - Homo sapiens.
OS Chimeric - Homo sapiens.
OS Synthetic.
PN WO9704092-A1.
PD 06-FEB-1997.
PE 17-JUL-1996; F01111.
PR 19-JUL-1995; FR-008729.
PA (RHON ) RHONE POULENC RORER SA.
PI Bracco L, Conseiller E;
DR WPI; 97-132633/12.
DR N-PSDB; T86215.
PT New p53 variants e.g. with oligomerisation domain replaced by
PT leucine zipper - useful for treating hyper-proliferative disorders,
PT esp. cancer and restenosis
PS Claim 30; Pages 76-78; 133pp; French.
CC Claimed variants of protein p53 have at least part of the
CC oligomerisation domain deleted and replaced by a leucine zipper
CC domain. The mutants preferably also have at least part of the p53
CC transactivation domain (amino acids 1-74) deleted and replaced by
CC the transactivating domain (TD) from herpes simplex virus viral
CC protein VP16 (amino acids 411-490). The present sequence is that of
CC a specifically claimed p53 variant designated V-325 and comprising
CC the VP16 TD, amino acids 75-325 of human wild-type p53 and a
CC leucine zipper domain at the C-terminal. The p53 variants are
CC more active and more stable tumour suppressors and apoptosis-inducing
```

CC agents than wild-type p53 and are active where the wild-type protein  
 CC is not, i.e. they are not inactivated by dominant negative or oncogenic  
 CC mutants, nor by other cellular proteins (because the leucine zipper  
 CC domain prevents formation of inactive mixed oligomers).  
 SQ Sequence 363 AA;

Query Match 100.0%; Score 86; DB 1; Length 363;  
 Best Local Similarity 100.0%; Pred. No. 5.46e-03;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 135 SPALNKMFCOL 145  
 |||||  
 QY 1 SPALNKMFCOL 11

## RESULT 14

ID W13954 standard; Protein; 363 AA.  
 AC W13954;  
 DT 25-JUN-1997 (first entry)  
 DE Modified p53 variant (del364-393).  
 KW p53; tumour suppressor; cancer; therapy; cell proliferation;  
 KW apoptosis; protein engineering; DNA binding.  
 OS Synthetic.  
 PN WO9710843-A1.  
 PD 27-MAR-1997.  
 PF 20-SEP-1996; U15188.  
 PR 22-SEP-1995; US-004802.  
 PR 21-AUG-1996; US-697221.  
 PA (WIST-) WISTAR INST ANATOMY & BIOLOGY.  
 PI Halazonetis TD;  
 PT WPI; 97-202618/18.  
 PT R284K modified p53 protein having DNA binding ability - useful in  
 treatment of cancer  
 PS Example 1: 49-51: 82pp; English.  
 CC A modified p53 variant (W13954) comprises wild-type p53 (see  
 CC also W13948) having a deletion of the C-terminal 30 amino acids,  
 CC and is obtd. by site-directed mutagenesis of p53 DNA. Deletion of  
 CC the p53 C-terminal 30 amino acids activates the DNA binding of  
 CC common class I p53 mutants (see also W13951-52). Novel modified  
 CC p53 variants (W13949-50, W13953-54, W13968-77), some contg.  
 CC C-terminal deletions, provide the means for pharmacological rescue  
 CC of p53 function in cancer patients. Nucleic acids coding for  
 CC modified p53 can be used for cancer gene therapy.  
 SQ Sequence 363 AA;

Query Match 100.0%; Score 86; DB 1; Length 363;  
 Best Local Similarity 100.0%; Pred. No. 5.46e-03;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 127 SPALNKMFCOL 137  
 |||||  
 QY 1 SPALNKMFCOL 11

## RESULT 15

ID W13972 standard; Protein; 363 AA.  
 AC W13972;  
 DT 25-JUN-1997 (first entry)  
 DE Modified p53 variant p53Q248del364-393.  
 KW p53; tumour suppressor; cancer; therapy; cell proliferation;  
 KW apoptosis; protein engineering; DNA binding.  
 OS Synthetic.  
 PN WO9710843-A1.  
 PD 27-MAR-1997.  
 PF 20-SEP-1996; U15188.  
 PR 22-SEP-1995; US-004802.  
 PR 21-AUG-1996; US-697221.  
 PA (WIST-) WISTAR INST ANATOMY & BIOLOGY.  
 PI Halazonetis TD;  
 PT WPI; 97-202618/18.  
 PT R284K modified p53 protein having DNA binding ability - useful in  
 treatment of cancer.  
 PS Example 1: 53-54: 82pp; English.

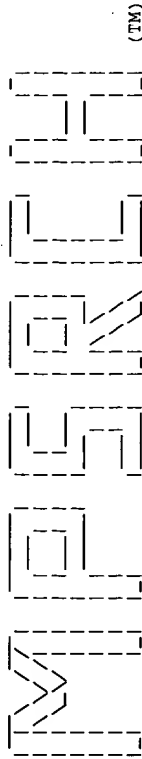
CC Modified p53 variant p53Q248del364-393 (W13972) has the tumour-  
 CC derived glutamine 248 mutation (see also W13951) and a deletion  
 CC of the C-terminal 30 amino acids of wild-type p53 (see also  
 CC W13948). Gln248 is a Class I p53 tumour mutation that affects DNA  
 CC binding. The C-terminal deletion, introduced by site-directed  
 CC mutagenesis of p53 DNA, activates the DNA binding of the p53  
 CC tumour mutant. This provides the means for pharmacological rescue  
 CC of p53 function in cancer patients. Other modified p53 constructs  
 CC (W13949-50, W13953-54, W13968-77) have also been produced. Nucleic  
 CC acids coding for modified p53 can be used for cancer gene therapy.  
 SQ Sequence 363 AA;

Query Match 100.0%; Score 86; DB 1; Length 363;  
 Best Local Similarity 100.0%; Pred. No. 5.46e-03;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 127 SPALNKMFCOL 137  
 |||||  
 QY 1 SPALNKMFCOL 11

Search completed: Sat Apr 15 00:28:26 2000  
 Job time : 35 secs.

\*\*\*\*\*



(TM)

\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Sat Apr 15 00:28:43 2000; MasPar time 3.25 Seconds  
Tabular output not generated. 135.622 Million cell updates/sec

Title: >US-08-452-843-15  
Description: (1-11) from US08452843.pep  
Perfect Score: 86  
Sequence: 1 SPALNKMFCQL 11

Scoring table: PAM 150  
Gap 15

Searched: 122810 seqs, 40068593 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: pir62  
1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 24.356; Variance 32.062; scale 0.760

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	86	100.0	393	1	cellular tumor antige	2.75e-07
2	82	95.3	393	2	cellular tumor antige	2.54e-06
3	77	89.5	386	2	cellular tumor antige	3.85e-05
4	77	89.5	393	2	tumor suppressor prot	3.85e-05
5	77	89.5	396	2	cellular tumor antige	3.85e-05
6	76	88.4	390	1	cellular tumor antige	6.59e-05
7	75	87.2	363	2	cellular tumor antige	1.12e-04
8	75	87.2	396	2	cellular tumor antige	1.12e-04
9	71	82.5	391	2	tumor suppressor p53	9.20e-04
10	65	75.6	391	2	cellular tumor antige	1.95e-02
11	63	73.3	367	2	cellular tumor antige	5.24e-02
12	56	65.1	381	2	cellular tumor antige	1.45e+00
13	55	64.0	1839	1	genome polyprotein -	2.29e+00
14	54	62.8	162	2	probable membrane pro	3.60e+00
15	54	62.8	344	2	GTP cyclohydrolase II	3.60e+00
16	54	62.8	344	2	GTP cyclohydrolase II	3.60e+00
17	53	61.6	421	2	hypothetical protein	5.62e+00
18	52	60.5	141	1	hemoglobin alpha chai	8.72e+00
19	52	60.5	141	1	hemoglobin alpha chai	8.72e+00
20	52	60.5	141	1	hemoglobin alpha chai	8.72e+00
21	52	60.5	141	1	hemoglobin alpha chai	8.72e+00
22	51	59.3	390	2	probable membrane pro	1.34e+01
23	51	59.3	470	2	hypothetical protein	1.34e+01

24 51 59.3 475 1 WZBEM4  
25 51 59.3 671 1 VCMVCE  
26 50 58.1 106 2 S20553  
27 50 58.1 111 1 A29654  
28 50 58.1 176 2 F69370  
29 50 58.1 259 1 WMBES2  
30 50 58.1 353 1 WMNV49  
31 50 58.1 353 2 C44221  
32 50 58.1 631 1 A48346  
33 50 58.1 631 1 VGN2PD  
34 50 58.1 855 2 S47533  
35 50 58.1 910 2 S40539  
36 50 58.1 1252 2 D71810  
37 50 58.1 1874 1 JQ0533  
38 49 57.0 119 2 JQ2032  
39 49 57.0 141 1 HANER  
40 49 57.0 378 2 S33994  
41 49 57.0 404 2 H70620  
42 49 57.0 524 2 T03112  
43 49 57.0 566 2 S37347  
44 49 57.0 587 2 E65171  
45 49 57.0 878 2 A55201

ALIGNMENTS

RESULT 1  
ENTRY DNH53 #type complete  
TITLE cellular tumor antigen p53 - human  
ALTERNATE\_NAMES cellular phosphoprotein p53; oncoprotein p53; transformation  
ORGANISM suppressor p53; tumor suppressor p53  
#formal\_name Homo sapiens #common\_name man  
DATE 05-Oct-1988 #sequence\_revision 18-Nov-1994 #text\_change  
26-Feb-1999  
ACCESSIONS A25224; A43073; J0436; S42659; A22837; A55060;  
A25397; B25397; S42452; S42453; I38082; I38083; I38084;  
I38085; I38086; I38087; I38089; I38090; I38091;  
I38092; I38093; A44905; I58354; I78850; I52681; S60153  
A25224  
REFERENCE  
#authors Lamb, P.; Crawford, L.  
#journal Mol. Cell. Biol. (1986) 6:1379-1385  
#title Characterization of the human p53 gene.  
#cross-references MIM:87064416  
#accession A25224  
#molecule\_type DNA  
#residues 1-393 #label LAM  
#cross-references EMBL:X01405; GB:M13121; GB:N00032; NID:g189460;  
PID:g386994  
REFERENCE J0436  
#authors Buchman, V.L.; Chumakov, P.M.; Ninkina, N.N.; Samarina, O.P.;  
Georgiev, G.P.  
#journal Gene (1988) 70:245-252  
#title A variation in the structure of the protein-coding region of  
the human p53 gene.  
#cross-references MIM:89108008  
#accession A43073  
#molecule\_type DNA  
#residues 1-393 #label BUC1  
#cross-references EMBL:M2898; NID:g189474  
#note this 72-Arg allele appears to be about 5 times more  
frequent than the 72-Pro allele  
#accession J0436  
#molecule\_type DNA  
#residues 1-71, 'P', 73-393 #label BUC2  
#cross-references EMBL:M2898; NID:g189474; PID:g189476  
#note this 72-Pro allele was found in both normal and  
malignant cell lines  
REFERENCE S40773  
#authors Chumakov, P.M.; Almazov, V.P.; Jenkins, J.R.  
#submission submitted to the EMBL Data Library, August 1990  
#accession S40773  
#molecule\_type DNA  
#residues 1-393 #label CHU

gene 17 protein - sai 1.34e+01  
env polyprotein - fel 1.34e+01  
cobyrinic acid a,c-di 2.06e+01  
proteinase inhibitor 2.06e+01  
conserved hypotheticala 2.06e+01  
28K protein - equine 2.06e+01  
orf3 protein - Autogr 2.06e+01  
cell fusion glycoprot 2.06e+01  
cell fusion glycoprot 2.06e+01  
glucose-6-phosphate 1 2.06e+01  
glucose-6-phosphate 1 2.06e+01  
probable type II DNA 2.06e+01  
genome polyprotein - 2.06e+01  
lambda 208 protein - 3.14e+01  
hemoglobin alpha chai 3.14e+01  
finger protein ZNF118 3.14e+01  
probable argj protein 3.14e+01  
minor capsid scaffold 3.14e+01  
syrd protein - Pseudo 3.14e+01  
hypothetical 64.0 kD 3.14e+01  
meiosis-specific prot 3.14e+01

##cross-references EMBL:X54156; NID:g35213; PID:g35214  
REFERENCE S42669  
#authors Matlashewski, G.; Lamb, P.; Pim, D.; Peacock, J.; Crawford, L.; Benchimol, S.  
#journal EMBO J. (1984) 3:3257-3262  
#title Isolation and characterization of a human p53 cDNA clone: expression of the human p53 gene.  
#cross-references MUID:85126934  
#accession S42669  
##molecule\_type mRNA  
##residues 101-393 ##label MKI1  
##cross-references EMBL:X01405; NID:g35215; PID:g642241  
REFERENCE A22837  
#authors Zakut-Houri, R.; Bienz-Tadmor, B.; Givol, D.; Oren, M.  
#journal EMBO J. (1985) 4:1251-1255  
#title Human p53 cellular tumor antigen: cDNA sequence and expression in COS cells.  
#cross-references MUID:85230377  
#accession A22837  
##molecule\_type mRNA  
##residues 1-71,'P',73-393 ##label ZAK  
##cross-references EMBL:X02469; EMBL:M50950; NID:g35209; PID:g35210  
REFERENCE A5060  
#authors Harlow, E.; Williamson, N.M.; Ralston, R.; Helfman, D.M.; Adams, T.E.  
#journal Mol. Cell. Biol. (1985) 5:1601-1610  
#title Molecular cloning and in vitro expression of a cDNA clone for human cellular tumor antigen p53.  
#cross-references MUID:85267676  
#accession A5060  
##molecule\_type mRNA  
##residues 1-71,'P',73-272,'H',274-393 ##label HAR  
##cross-references GB:X03199; NID:g189478; PID:g189479  
##experimental\_source clone pR4-2, cell line A431  
REFERENCE A93086  
#authors Harris, N.; Brill, E.; Shohat, O.; Prokocimer, M.; Wolf, D.; Aral, N.; Rotter, V.  
#journal Mol. Cell. Biol. (1986) 6:4650-4656  
#title Molecular basis for heterogeneity of the human p53 protein.  
#cross-references MUID:87089826  
#accession A25397  
##molecule\_type mRNA  
##residues 1-78,'T',80-393 ##label HAR1  
##cross-references EMBL:M14694; NID:g339813; PID:g339814  
##experimental\_source clone p53-H-1, transformed hybridoma SV-80 cell line  
#accession B25397  
##molecule\_type mRNA  
##residues 1-71,'P',73-78,'T',80-393 ##label HAR2  
##cross-references EMBL:M14695; NID:g339815; PID:g339816  
##experimental\_source clone p53-H-19, transformed hybridoma SV-80 cell line  
#accession S42452  
#authors Matlashewski, G.J.; Tuck, S.; Pim, D.; Lamb, P.; Schneider, J.; Crawford, L.V.  
#journal Mol. Cell. Biol. (1987) 7:961-963  
#title Primary structure polymorphism at amino acid residue 72 of human p53.  
#cross-references MUID:87144273  
#accession S42452  
##molecule\_type mRNA; DNA  
##residues 66-71,'P',73-79 ##label MKI2  
##experimental\_source clone lambda C113  
##note 72-Cys was also found, and appears to represent a polymorphism  
#accession S42453  
##molecule\_type mRNA; DNA  
##residues 66-79 ##label MKI3  
##experimental\_source clone J6K  
REFERENCE I38082  
#authors Farrell, P.J.; Allan, G.J.; Shanahan, F.; Vousden, K.H.; Crook, T.  
#journal EMBO J. (1991) 10:2879-2887

#title p53 is frequently mutated in Burkitt's lymphoma cell lines.  
#cross-references MUID:92007731  
#accession I38082  
##status translated from GB/EMBL/DBJ  
##molecule\_type mRNA  
##residues 1-189,'LLSILSEWKEICVSIWMTETLFDIYVWCPMSRLRLALT',  
'VPPSTTTCTVTPANAA' ##label F01  
##cross-references EMBL:X60010; NID:g506432; PID:g506433  
##note deletion of a C nucleotide causes a frameshift at position 566  
#accession I38083  
##status translated from GB/EMBL/DBJ  
##molecule\_type mRNA  
##residues 1-192,'R',194-393 ##label F02  
##cross-references EMBL:X60011; NID:g506434; PID:g506435  
#accession I38084  
##status translated from GB/EMBL/DBJ  
##molecule\_type mRNA  
##residues 1-393 ##label F03  
##cross-references EMBL:X60012; NID:g506436; PID:g506437  
#accession I38085  
##status translated from GB/EMBL/DBJ  
##molecule\_type mRNA  
##residues 1-245,'T',247-393 ##label F04  
##cross-references EMBL:X60013; NID:g506438; PID:g506439  
#accession I38086  
##status translated from GB/EMBL/DBJ  
##molecule\_type mRNA  
##residues 1-236,'I',238-393 ##label F05  
##cross-references EMBL:X60014; NID:g506440; PID:g506441  
#accession I38087  
##status translated from GB/EMBL/DBJ  
##molecule\_type mRNA  
##residues 1-247,'Q',249-393 ##label F06  
##cross-references EMBL:X60015; NID:g506442; PID:g506443  
#accession I38088  
##status translated from GB/EMBL/DBJ  
##molecule\_type mRNA  
##residues 1-71,'P',73-237,'Y',239-393 ##label F07  
##cross-references EMBL:X60016; NID:g506444; PID:g506445  
#accession I38089  
##status translated from GB/EMBL/DBJ  
##molecule\_type mRNA  
##residues 1-247,'Q',249-393 ##label F08  
##cross-references EMBL:X60017; NID:g506446; PID:g506447  
#accession I38090  
##status translated from GB/EMBL/DBJ  
##molecule\_type mRNA  
##residues 1-71,'P',73-162,'H',164-393 ##label F09  
##cross-references EMBL:X60018; NID:g506448; PID:g506449  
#accession I38091  
##status translated from GB/EMBL/DBJ  
##molecule\_type mRNA  
##residues 1-212,'Q',214-393 ##label F10  
##cross-references EMBL:X60019; NID:g506450; PID:g506451  
#accession I38092  
##status translated from GB/EMBL/DBJ  
##molecule\_type mRNA  
##residues 1-253,'D',255-393 ##label F11  
##cross-references EMBL:X60020; NID:g506452; PID:g506453  
##note all sequences submitted to the EMBL/GenBank/DBJ databases June 1991  
REFERENCE I38093  
#authors Futreal, P.A.; Barrett, J.C.; Wiseman, R.W.  
#journal Nucleic Acids Res. (1991) 19:6977  
#title An Alu polymorphism intragenic to the TP53 gene.  
#cross-references MUID:92107726  
#accession I38093  
##status translated from GB/EMBL/DBJ  
##molecule\_type DNA  
##residues 1-393 ##label FUT  
##cross-references EMBL:X54156; NID:g35213; PID:g35214  
REFERENCE A44905

...  
Note: remainder of annotations omitted.

```

RESULT      5
ENTRY
TITLE      JH0633      #type complete
ALTERNATE_NAMES      cellular tumor antigen p53 - golden hamster
ORGANISM      tumor-suppressor protein p53
ORGANISM      #formal_name Mesocricetus auratus #common_name golden hamster
DATE      17-Aug-1992 #sequence_revision 17-Aug-1992 #text_change
.
ACCESSIONS      JH0633
REFERENCE      JH0633
#authors      Legros, Y.; McIntyre, P.; Soussi, T.
#journal      Gene (1992) 112:247-250
#title      The cDNA cloning and immunological characterization of
              hamster p53.
#cross-references      MUID:92210007

```

```
#accession JH0633
#molecule_type mRNA
#residues 1-396 ##label LEG
##cross-references GB:M75144; NID:g191414; PID:g191415
##experimental_source kidney, strain MP1
GENETICS
#gene p53
CLASSIFICATION
#superfamily cellular tumor antigen p53
KEYWORDS
#apoptosis; cell division control; DNA binding; homotetramer;
#nucleus; phosphoprotein; transcription regulation; tumor
#suppressor; zinc
FEATURE
179,182,241,245 #binding_site zinc (Cys, His, Cys, Cys) #status
#predicted\
395 #binding_site phosphoryl-RNA (Ser) (covalent) #status
#predicted\
SUMMARY
#length 396 #molecular-weight 43631 #checksum 6617
Query Match 89.5%; Score 77; DB 2; Length 396;
Best Local Similarity 81.8%; Pred. No. 3.85e-05;
Matches 9; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
Db 130 SPLNKLFCOL 140
||||:||||
QY 1 SPALNKMFCOL 11
RESULT 6
ENTRY ##type complete
TITLE cellular tumor antigen p53 - mouse
ALTERNATE_NAMES oncoprotein p53
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE 28-Aug-1985 #sequence_revision 04-Oct-1996 #text_change
12-Feb-1999
ACCESSIONS A22739; S06336; A02684; S38822; S40014; I48703
REFERENCE A22739
#authors Bienz, B.; Zakut-Houri, R.; Givol, D.; Oren, M.
#journal EMBO J. (1984) 3:2179-2183
#cross-references MUID:85027173
#accession A22739
#molecule_type DNA
#residues 1-134,'V',136-390 ##label BIE
##cross-references GB:X00876; NID:g871420; PID:g871421; GB:X01237;
GB:K01700; NID:g53575; PID:g53576
REFERENCE S06336
#authors Chumakov, P.M.
#journal Bioorg. Khim. (1987) 13:1691-1694
#title Primary structure of DNA complementary to murine oncoprotein
p53 mRNA.
#cross-references MUID:88221682
#accession S06336
#status not compared with conceptual translation
#molecule_type mRNA
#residues 1-134,'V',136-390 ##label CHU
REFERENCE A02684
#authors Zakut-Houri, R.; Oren, M.; Bienz, B.; Lavie, V.; Hazum, S.;
Givol, D.
#journal Nature (1983) 306:594-597
#title A single gene and a pseudogene for the cellular tumour
antigen p53.
#cross-references MUID:84068204
#accession A02684
#molecule_type mRNA
#residues 1-159,'H',161-167,'G',169-233,'I',235-390 ##label ZAK
##cross-references GB:X01237; GB:K01700; NID:g53575
REFERENCE S38822
#authors Arai, N.; Nomura, D.; Yokota, K.; Wolf, D.; Brill, E.;
Shohat, O.; Rotter, V.
#journal Mol. Cell. Biol. (1986) 6:3232-3239
#title Immunologically distinct p53 molecules generated by
alternative splicing.
#cross-references MUID:87064640
#accession S38822
```

```
##status preliminary
##molecule_type mRNA
#residues 1-390 ##label ARA1
##cross-references EMBL:M13872; NID:g200198; PID:g200199
#accession S38823
#status preliminary
##molecule_type mRNA
#residues 1-167,'G',169-233,'I',235-390 ##label ARA2
##cross-references EMBL:M13873
REFERENCE S40014
#authors Arai, N.; Nomura, D.; Yokota, K.; Wolf, D.; Brill, E.;
Shohat, O.; Rotter, V.
#submission submitted to the EMBL Data Library, July 1988
#accession S40014
##molecule_type mRNA
#residues 1-167,'G',169-390 ##label ARA3
##cross-references EMBL:M13873; NID:g200200; PID:g200201
REFERENCE I48703
#authors Jenkins, J.R.; Rudge, K.; Redmond, S.; Wade-Evans, A.
#journal Nucleic Acids Res. (1984) 12:5609-5626
#title Cloning and expression analysis of full length mouse cDNA
sequences encoding the transformation associated protein
p53.
#cross-references MUID:84272240
#accession I48703
#status preliminary; translated from GB/EMBL/DDBJ
##molecule_type mRNA
#residues 1-47,'R',49-78,'QW',82-390 ##label RES
##cross-references EMBL:X00741; NID:g53570; PID:g53571
COMMENT This DNA-binding protein plays an essential role in the regulation
of cell division, as it is required for the transition from phase
G0 to G1 of the cell cycle.
COMMENT The tetramer association region may exhibit a beta-turn,
beta-sheet, beta-turn, alpha-helix motif.
CLASSIFICATION #superfamily cellular tumor antigen p53
KEYWORDS apoptosis; cell division control; DNA binding; homotetramer;
phosphoprotein; transcription regulation; tumor suppressor;
zinc
FEATURE
1-44 #domain transcription activation #status predicted
#label TRA\
16-26 #region conserved region I\
99-289 #domain DNA-binding core #status predicted #label DBC\
108-121 #region L1 loop\
114-139 #region conserved region II\
160-192 #region L2 loop\
168-178 #region conserved region III\
231-232 #region conserved region IV\
233-248 #region L3 loop\
267-283 #region conserved region V\
313-319 #region nuclear location signal\
315-357 #region tetramer association\
7,9,12,18,23,37 #binding_site phosphate (Ser) (covalent) #status
predicted\
173,176,235,239 #binding_site zinc (Cys, His, Cys, Cys) #status
predicted\
312 #binding_site phosphate (Ser) (covalent) (by cdc2
kinase) #status predicted\
389 #binding_site phosphoryl-RNA (Ser) (covalent) #status
predicted\
SUMMARY #length 390 #molecular-weight 43458 #checksum 1260
Query Match 88.4%; Score 76; DB 1; Length 390;
Best Local Similarity 81.8%; Pred. No. 6.59e-05;
Matches 9; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
Db 124 SPLNKLFCOL 134
||||:||||
QY 1 SPALNKMFCOL 11
RESULT 7
ENTRY ##type complete
A29376
```



#authors Yamada, Y.; Yoshida, T.; Hayashi, K.; Sekiya, T.; Yokota, J.; Hirohashi, S.; Nakatani, K.; Nakano, H.; Sugimura, T.; Terada, M.  
#journal Cancer Res. (1991) 51:5800-5805  
#title p53 gene mutations in gastric cancer metastases and in gastric cancer cell lines derived from metastases.  
#cross-references MUID:92034678  
#accession A44905

...  
Note: remainder of annotations omitted.

Query Match: 100.0%; Score 86; DB 1; Length 393;  
Best Local Similarity 100.0%; Pred. No. 2.75e-07;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 127 SPALNKMFCQL 137  
| | | | | | | | | |  
Qy 1 SPALNKMFCQL 11

RESULT 2  
ENTRY #type complete  
TITLE cellular tumor antigen p53 - green monkey  
ORGANISM #formal\_name Cercopithecus aethiops #common\_name green monkey, grivet  
DATE 28-Feb-1990 #sequence\_revision 28-Feb-1990 #text\_change

ACCESSIONS S06594  
REFERENCE S06594  
#authors Rigaudy, P.; Eckhart, W.  
#journal Nucleic Acids Res. (1989) 17:8375  
#title Nucleotide sequence of a cDNA encoding the monkey cellular phosphoprotein p53.  
#cross-references MUID:90045967  
#accession S06594

#molecule\_type mRNA  
#residues 1-393 #label RIG  
#cross-references EMBL:X16384; NID:g22795; PID:g22796  
CLASSIFICATION #superfamily cellular tumor antigen p53  
KEYWORDS apoptosis; cell division control; DNA binding; homotetramer; nucleus; phosphoprotein; transcription regulation; tumor suppressor; zinc

FEATURE 176,179,238,242 #binding\_site zinc (Cys, His, Cys) #status predicted  
392 #binding\_site phosphoryl-RNA (Ser) (covalent) #status predicted  
SUMMARY #length 393 #molecular-weight 43696 #checksum 4263

Query Match 95.3%; Score 82; DB 2; Length 393;  
Best Local Similarity 90.9%; Pred. No. 2.54e-06;  
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 127 SPDLNKMFCQL 137  
| | | | | | | | | |  
Qy 1 SPALNKMFCQL 11

RESULT 3  
ENTRY #type complete  
TITLE cellular tumor antigen p53 - bovine  
ALTERNATE\_NAMES tumor-suppressor protein p53  
ORGANISM #formal\_name Bos primigenius taurus #common\_name cattle  
DATE 07-May-1995 #sequence\_revision 01-Sep-1995 #text\_change

ACCESSIONS S51648  
REFERENCE S51648  
#authors Dequiedt, F.; Willems, L.; Burny, A.; Kettmann, R.  
#submission submitted to the EMBL Data Library, September 1994  
#description Nucleotide sequence of the ovine p53 tumor-suppressor gene cDNA and its genomic organisation.

#accession S51648  
#status preliminary

#molecule\_type mRNA  
#residues 1-386 #label DEQ  
#cross-references EMBL:X81704; NID:g602332; PID:g602333  
CLASSIFICATION #superfamily cellular tumor antigen p53  
KEYWORDS apoptosis; cell division control; DNA binding; homotetramer; phosphoprotein; transcription regulation; tumor suppressor; zinc

FEATURE 168,171,231,235 #binding\_site zinc (Cys, His, Cys) #status predicted  
385 #binding\_site phosphoryl-RNA (Ser) (covalent) #status predicted  
SUMMARY #length 386 #molecular-weight 43255 #checksum 7025

Query Match 89.5%; Score 77; DB 2; Length 386;  
Best Local Similarity 81.8%; Pred. No. 3.85e-05;  
Matches 9; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 119 SPSLNKLFCQL 129  
| | | | | | | | | |  
Qy 1 SPALNKMFCQL 11

RESULT 4  
ENTRY #type complete  
TITLE tumor suppressor protein p53 - Chinese hamster  
ORGANISM #formal\_name Cricetus griseus #common\_name Chinese hamster  
DATE 11-Apr-1997 #sequence\_revision 09-May-1997 #text\_change

ACCESSIONS JC6176  
REFERENCE JC6176  
#authors Lee, H.; Larner, J.M.; Hamlin, J.L.  
#journal Gene (1997) 184:177-183  
#title Cloning and characterization of Chinese hamster p53 cDNA.  
#cross-references MUID:97183659  
#contents liver  
#accession JC6176

#molecule\_type mRNA  
#residues 1-393 #label LEE  
#cross-references GB:U50395; NID:g1842229; PID:g1842230  
COMMENT This protein is a multimer. It plays the central role in a complex DNA damage-sensing network. It binds to replication factor and TATA-binding protein, and affects DNA replication, transcription, and recombination by protein/protein interactions.

GENETICS  
#gene p53  
CLASSIFICATION #superfamily cellular tumor antigen p53  
KEYWORDS liver; tumor  
SUMMARY #length 393 #molecular-weight 43362 #checksum 4043

Query Match 89.5%; Score 77; DB 2; Length 393;  
Best Local Similarity 81.8%; Pred. No. 3.85e-05;  
Matches 9; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 127 SPSLNKLFCQL 137  
| | | | | | | | | |  
Qy 1 SPALNKMFCQL 11

RESULT 5  
ENTRY #type complete  
TITLE cellular tumor antigen p53 - golden hamster  
ALTERNATE\_NAMES tumor-suppressor protein p53  
ORGANISM #formal\_name Mesocricetus auratus #common\_name golden hamster  
DATE 17-Aug-1992 #sequence\_revision 17-Aug-1992 #text\_change

ACCESSIONS JH0633  
REFERENCE JH0633  
#authors Legros, Y.; McIntyre, P.; Soussi, T.  
#journal Gene (1992) 112:247-250  
#title The cDNA cloning and immunological characterization of hamster p53.  
#cross-references MUID:92210007

```

#accession JH0633
#molecule_type mRNA
#residues 1-396 #label LEG
#cross-references GB:M75144; NID:g191414; PID:g191415
#experimental_source kidney, strain MP1
GENETICS
#gene p53
#superfamily cellular tumor antigen p53
CLASSIFICATION
#apoptosis; cell division control; DNA binding; homotetramer;
#nucleus; phosphoprotein; transcription regulation; tumor
#suppressor; zinc
KEYWORDS
FEATURE
179,182,241,245 #binding_site zinc (Cys, His, Cys, Cys) #status
predicted\
395 #binding_site phosphoryl-RNA (Ser) (covalent) #status
predicted\
SUMMARY
#length 396 #molecular-weight 43631 #checksum 6617
Query Match 89.5%; Score 77; DB 2; Length 396;
Best Local Similarity 81.8%; Pred. No. 3.85e-05;
Matches 9; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
Db 130 SPLNKLFCOL 140
||:|||||
QY 1 SPALNKMFCOL 11

RESULT 6
ENTRY DNMS53 #type complete
TITLE cellular tumor antigen p53 - mouse
ALTERNATE_NAMES oncoprotein p53
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE 28-Aug-1985 #sequence_revision 04-Oct-1996 #text_change
12-Feb-1999
ACCESSIONS A22739; S06336; A02684; S38822; S40014; I48703
REFERENCE A22739
#authors Blenz, B.; Zakut-Houri, R.; Givol, D.; Oren, M.
#journal EMBO J. (1984) 3:2179-2183
#cross-references MUID:85027173
#accession A22739
#molecule_type DNA
#residues 1-134, 'V', 136-390 #label BIE
#cross-references GB:X00876; NID:g871420; PID:g871421; GB:X01237;
GB:K01700; NID:g53575; PID:g53576
REFERENCE S06336
#authors Chumakov, P.M.
#journal Bioorg. Khim. (1987) 13:1691-1694
#title Primary structure of DNA complementary to murine oncoprotein
p53 mRNA.
#cross-references MUID:88221682
#accession S06336
#status not compared with conceptual translation
#molecule_type mRNA
#residues 1-134, 'V', 136-390 #label CHU
REFERENCE A02684
#authors Zakut-Houri, R.; Oren, M.; Blenz, B.; Lavie, V.; Hazum, S.;
Givol, D.
#journal Nature (1983) 306:594-597
#title A single gene and a pseudogene for the cellular tumour
antigen p53.
#cross-references MUID:84068204
#accession A02684
#molecule_type mRNA
#residues 1-159, 'H', 161-167, 'G', 169-233, 'I', 235-390 #label ZAK
#cross-references GB:X01237; GB:K01700; NID:g53575
REFERENCE S38822
#authors Arai, N.; Nomura, D.; Yokota, K.; Wolf, D.; Brill, E.;
Shohat, O.; Rotter, V.
#journal Mol. Cell. Biol. (1986) 6:3232-3239
#title Immunologically distinct p53 molecules generated by
alternative splicing.
#cross-references MUID:87064640
#accession S38822

```

```

##status preliminary
##molecule_type mRNA
#residues 1-390 #label ARA1
#cross-references EMBL:M13872; NID:g200198; PID:g200199
#accession S38823
##status preliminary
##molecule_type mRNA
#residues 1-167, 'G', 169-233, 'I', 235-390 #label ARA2
#cross-references EMBL:M13873
REFERENCE S40014
#authors Arai, N.; Nomura, D.; Yokota, K.; Wolf, D.; Brill, E.;
Shohat, O.; Rotter, V.
#submission submitted to the EMBL Data Library, July 1988
#accession S40014
##molecule_type mRNA
#residues 1-167, 'G', 169-390 #label ARA3
#cross-references EMBL:M13873; NID:g200200; PID:g200201
REFERENCE I48703
#authors Jenkins, J.R.; Rudge, K.; Redmond, S.; Wade-Evans, A.
#journal Nucleic Acids Res. (1984) 12:5609-5626
#title Cloning and expression analysis of full length mouse cDNA
sequences encoding the transformation associated protein
p53.
#cross-references MUID:84272240
#accession I48703
##status preliminary; translated from GB/EMBL/DDBJ
##molecule_type mRNA
#residues 1-47, 'R', 49-78, 'QW', 82-390 #label RES
#cross-references EMBL:X00741; NID:g53570; PID:g53571
COMMENT This DNA-binding protein plays an essential role in the regulation
of cell division, as it is required for the transition from phase
G0 to G1 of the cell cycle.
COMMENT The tetramer association region may exhibit a beta-turn,
beta-sheet, beta-turn, alpha-helix motif.
CLASSIFICATION #superfamily cellular tumor antigen p53
KEYWORDS apoptosis; cell division control; DNA binding; homotetramer;
phosphoprotein; transcription regulation; tumor suppressor;
zinc
FEATURE
1-44 #domain transcription activation #status predicted
#label TRA\
16-26 #region conserved region I\
99-289 #domain DNA-binding core #status predicted #label DBC\
108-121 #region L1 loop\
114-139 #region conserved region II\
160-192 #region L2 loop\
168-178 #region conserved region III\
231-252 #region conserved region IV\
233-248 #region L3 loop\
267-283 #region conserved region V\
313-319 #region nuclear location signal\
319-357 #region tetramer association\
7,9,12,18,23,37 #binding_site phosphate (Ser) (covalent) #status
predicted\
173,176,235,239 #binding_site zinc (Cys, His, Cys, Cys) #status
predicted\
312 #binding_site phosphate (Ser) (covalent) (by cdc2
kinase) #status predicted\
389 #binding_site phosphoryl-RNA (Ser) (covalent) #status
predicted\
SUMMARY #length 390 #molecular-weight 43458 #checksum 1260
Query Match 88.4%; Score 76; DB 1; Length 390;
Best Local Similarity 81.8%; Pred. No. 6.59e-05;
Matches 9; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
Db 124 SPLNKLFCOL 134
||:|||||
QY 1 SPALNKMFCOL 11

RESULT 7
ENTRY A29376 #type complete

```

```

TITLE      cellular tumor antigen p53 - African clawed frog
ORGANISM   #formal_name Xenopus laevis #common_name African clawed frog
DATE       31-Mar-1989 #sequence_revision 31-Mar-1989 #text_change
08-Sep-1997

ACCESSIONS A29376, S61531; S72313; I51639
REFERENCE   A29376
AUTHORS    Soussi, T.; de Fromental, C.C.; Mechali, M.; May, P.; Kress, M.
JOURNAL    Oncogene (1987) 1:71-78
TITLE      Cloning and characterization of a cDNA from Xenopus laevis
           coding for a protein homologous to human and murine p53.
CROSS-REFERENCES MUID:88143684
ACCESSION   A29376
MOLECULE_TYPE mRNA
RESIDUES    1-363 #label SOU
CROSS-REFERENCES EMBL:X05191; NID:g64961; PID:g64962
REFERENCE   I51639
AUTHORS    Hoever, M.; Clement, J.H.; Wedlich, D.; Montenarh, M.; Knochel, W.
JOURNAL    Oncogene (1994) 9:109-120
TITLE      Overexpression of wild-type p53 interferes with normal
           development in Xenopus laevis embryos.
CROSS-REFERENCES MUID:94134403
ACCESSION   S61531
MOLECULE_TYPE mRNA
RESIDUES    1-293,295-363 #label HOE
CROSS-REFERENCES EMBL:X77546; NID:g468513; PID:g468514
REFERENCE   S72313
AUTHORS    Hoever, M.; Clement, J.; Wedlich, D.; Montenarh, M.; Knochel, W.
SUBMISSION submitted to the EMBL Data Library, March 1994
ACCESSION   S72313
MOLECULE_TYPE mRNA
RESIDUES    1-51,'S',53-70,72-293,295-363 #label HOW
CROSS-REFERENCES EMBL:X77546; NID:g468513; PID:g468514
GENETICS
GENE        p53
CLASSIFICATION #superfamily cellular tumor antigen p53
KEYWORDS     apoptosis; cell division control; DNA binding; homotetramer;
           nucleus; phosphoprotein; transcription regulation; tumor
           suppressor; zinc
FEATURE
150,153,213,217 #binding_site zinc (Cys, His, Cys, Cys) #status
                predicted\
362             #binding_site phosphoryl-RNA (Ser) (covalent) #status
                predicted\
SUMMARY      #length 363 #molecular-weight 40692 #checksum 6648
Query Match 87.2%; Score 75; DB 2; Length 363;
Best Local Similarity 81.8%; Pred. No. 1.12e-04;
Matches 9; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
Db 101 SPALNKLFCQL 111
||| |||:||||
QY 1 SPALNKMFCQL 11
RESULT      8
ENTRY       JH0631 #type complete
TITLE      cellular tumor antigen p53 - rainbow trout
ORGANISM   #formal_name Oncorhynchus mykiss #common_name rainbow trout
DATE       17-Aug-1992 #sequence_revision 17-Aug-1992 #text_change
08-Sep-1997
ACCESSIONS JH0631
REFERENCE   JH0631
AUTHORS    de Fromental, C.C.; Pakdel, F.; Chapus, A.; Baney, C.; May, P.; Soussi, T.
JOURNAL    Gene (1992) 112:241-245
TITLE      Rainbow trout p53: cDNA cloning and biochemical
           characterization.
CROSS-REFERENCES MUID:92210006
ACCESSION   JH0631
MOLECULE_TYPE mRNA

```

```

RESIDUES    1-396 #label DEF
CROSS-REFERENCES GB:M75145; NID:g213828; PID:g213829
EXPERIMENTAL_source liver
COMMENT     This protein is the product of a tumor suppressor gene, p53, whose
           inactivation leads to cell transformation or neoplasia.
CLASSIFICATION #superfamily cellular tumor antigen p53
KEYWORDS     apoptosis; cell division control; DNA binding; homotetramer;
           nucleus; phosphoprotein; transcription regulation; tumor
           suppressor; zinc
FEATURE
164,167,227,231 #binding_site zinc (Cys, His, Cys, Cys) #status
                predicted\
395             #binding_site phosphoryl-RNA (Ser) (covalent) #status
                predicted\
SUMMARY      #length 396 #molecular-weight 43966 #checksum 9018
Query Match 87.2%; Score 75; DB 2; Length 396;
Best Local Similarity 81.8%; Pred. No. 1.12e-04;
Matches 9; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
Db 115 SPDLNKLFCQL 125
||| |||:||||
QY 1 SPALNKMFCQL 11
RESULT      9
ENTRY       JC6193 #type complete
TITLE      tumor suppressor p53 - rabbit
ORGANISM   #formal_name Oryctolagus cuniculus #common_name domestic
           rabbit
DATE       11-Apr-1997 #sequence_revision 09-May-1997 #text_change
17-Mar-1999
ACCESSIONS JC6193
REFERENCE   JC6193
AUTHORS    Le Goas, F.; May, P.; Ronco, P.; de Fromental, C.C.
JOURNAL    Gene (1997) 185:169-173
TITLE      cDNA cloning and immunological characterization of rabbit
           p53.
CROSS-REFERENCES MUID:97208869
ACCESSION   JC6193
MOLECULE_TYPE mRNA
RESIDUES    1-391 #label LEA
CROSS-REFERENCES EMBL:X90592; NID:gl532043; PID:el94962; PID:gl532044
GENETICS
GENE        p53
CLASSIFICATION #superfamily cellular tumor antigen p53
KEYWORDS     tumor
SUMMARY      #length 391 #molecular-weight 43435 #checksum 4367
Query Match 82.6%; Score 71; DB 2; Length 391;
Best Local Similarity 81.8%; Pred. No. 9.20e-04;
Matches 9; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
Db 124 SPCLNKLFCQL 134
||| |||:||||
QY 1 SPALNKMFCQL 11
RESULT      10
ENTRY       S02192 #type complete
TITLE      cellular tumor antigen p53 - rat
ORGANISM   #formal_name Rattus norvegicus #common_name Norway rat
DATE       18-Oct-1989 #sequence_revision 18-Oct-1989 #text_change
17-Mar-1999
ACCESSIONS S02192; S41149
REFERENCE   S02192
AUTHORS    Soussi, T.; de Fromental, C.C.; Breugnot, C.; May, E.
JOURNAL    Nucleic Acids Res. (1988) 16:11384
TITLE      Nucleotide sequence of a cDNA encoding the rat p53 nuclear
           oncoprotein.
CROSS-REFERENCES MUID:89083585
ACCESSION   S02192

```

```
##molecule_type mrna
##residues 1-391 ##label SOU
##cross-references EMBL:X13058; NID:g56828; PID:g56829
REFERENCE S41149
#authors Hulla, J.E.; Schneider, R.P.
#journal Nucleic Acids Res. (1993) 21:713-717
#title Structure of the rat p53 tumor suppressor gene.
#cross-references MUID:93181268
#accession S41149
#status Preliminary; nucleic acid sequence not shown;
translation not shown
##molecule_type DNA
##residues 1-173,'W',175-391 ##label HUL
##cross-references EMBL:L07909
##note the nucleotide sequence was submitted to the EMBL Data
Library, December 1992
GENETICS 25/2; 32/3; 123/3; 185/1; 259/2; 305/1; 329/3; 365/2
CLASSIFICATION #superfamily cellular tumor antigen p53
KEYWORDS apoptosis; cell division control; DNA binding; homotetramer;
nucleus; phosphoprotein; transcription regulation; tumor
suppressor; zinc
FEATURE 174,177,236,240 #binding_site zinc (Cys, His, Cys, Cys) #status
predicted\
390 #binding_site phosphoryl-RNA (Ser) (covalent) #status
predicted\
SUMMARY #length 391 #molecular-weight 43451 #checksum 7105
Query Match 75.6%; Score 65; DB 2; Length 391;
Best Local Similarity 77.8%; Pred. No. 1.95e-02;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
Db 127 SLNKLFCQL 135
:|||||
QY 3 ALNKMFCQL 11
RESULT 11
ENTRY S02193 #type complete
TITLE cellular tumor antigen p53 - chicken
ALTERNATE_NAMES nuclear oncoprotein p53
ORGANISM #formal_name Gallus gallus #common_name chicken
DATE 18-Oct-1989 #sequence_revision 18-Oct-1989 #text_change
08-Sep-1997
ACCESSIONS S02193
REFERENCE S02193
#authors Soussi, T.; Begue, A.; Kress, M.; Stehelin, D.; May, P.
#journal Nucleic Acids Res. (1988) 16:11383
#title Nucleotide sequence of a cDNA encoding the chicken p53
nuclear oncoprotein.
#cross-references MUID:89083584
#accession S02193
##molecule_type mrna
##residues 1-367 ##label SOU
##cross-references EMBL:X13057; NID:g63740; PID:g63741
CLASSIFICATION #superfamily cellular tumor antigen p53
KEYWORDS apoptosis; cell division control; DNA binding; homotetramer;
nucleus; phosphoprotein; transcription regulation; tumor
suppressor; zinc
FEATURE 161,164,224,228 #binding_site zinc (Cys, His, Cys, Cys) #status
predicted\
366 #binding_site phosphoryl-RNA (Ser) (covalent) #status
predicted\
SUMMARY #length 367 #molecular-weight 40169 #checksum 5094
Query Match 73.3%; Score 63; DB 2; Length 367;
Best Local Similarity 63.6%; Pred. No. 5.24e-02;
Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
Db 112 SPVLNKKVQRL 122
|||||
```

```
QY 1 SPALNKMFCQL 11
RESULT 12
ENTRY S38824 #type complete
TITLE cellular tumor antigen p53, minor splice form - mouse
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change
17-Mar-1999
ACCESSIONS S38824; S35478
REFERENCE S38822
#authors Arai, N.; Nomura, D.; Yokota, K.; Wolf, D.; Brill, E.;
Shohat, O.; Rotter, V.
#journal Mol. Cell. Biol. (1986) 6:3232-3239
#title Immunologically distinct p53 molecules generated by
alternative splicing.
#cross-references MUID:87064640
#accession S38824
##molecule_type mrna
##residues 1-381 ##label ARA
##cross-references GB:M13874; NID:g200202; PID:g200203
REFERENCE S35478
#authors Han, K.A.; Kulesz-Martin, M.F.
#journal Nucleic Acids Res. (1992) 20:1979-1981
#title Alternatively spliced p53 RNA in transformed and normal cells
of different tissue types.
#cross-references MUID:92253421
#accession S35478
##status nucleic acid sequence not shown; translation not shown
##molecule_type mRNA
##residues 1-381 ##label HAN
##cross-references EMBL:M13874; NID:g200202; PID:g200203
##note the nucleotide sequence was submitted to the EMBL Data
Library, July 1988
COMMENT This sequence, produced by alternative splicing of the tenth
intron, lacks the carboxyl-terminal sequence necessary for
covalent attachment of RNA. The function of this minor splice
form is not known.
CLASSIFICATION #superfamily cellular tumor antigen p53
KEYWORDS alternative splicing; phosphoprotein; zinc
FEATURE 1-44
#domain transcription activation #status predicted
16-26 #label TRA\
99-289 #region conserved region I\
108-121 #domain DNA-binding core #status predicted #label DBC\
114-139 #region L1 loop\
160-192 #region conserved region II\
168-178 #region L2 loop\
231-252 #region conserved region III\
233-248 #region conserved region IV\
267-283 #region L3 loop\
313-319 #region conserved region V\
319-357 #region nuclear location signal\
7,9,12,18,23,37 #region tetramer association\
173,176,235,239 #binding_site phosphate (Ser) (covalent) #status
predicted\
312 #binding_site zinc (Cys, His, Cys, Cys) #status
predicted\
#binding_site phosphate (Ser) (covalent) (by cdc2
kinase) #status predicted
SUMMARY #length 381 #molecular-weight 42498 #checksum 8703
Query Match 65.1%; Score 56; DB 2; Length 381;
Best Local Similarity 72.7%; Pred. No. 1.45e+00;
Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Db 124 SPPLNKLFFQL 134
|||||
QY 1 SPALNKMFCQL 11
RESULT 13
ENTRY RRPPEM #type complete
```

```

TITLE      genome polypeptide - eggplant mosaic virus
ALTERNATE_NAMES  RNA nucleotidyltransferase (RNA-directed); RNA replicase
CONTAINS     RNA-directed RNA polymerase (EC 2.7.7.48)
ORGANISM     #formal_name eggplant mosaic virus
DATE         30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change
ACCESSIONS   JQ0102
REFERENCE     Osorio-Keese, M.E.; Keese, P.; Gibbs, A.
#authors     Virology (1989) 172:547-554
#journal     Nucleotide sequence of the genome of eggplant mosaic
#title       tymovirus.
#cross-references MUID:90021185
#accession   JQ0102
#molecule_type genomic RNA
##residues  1-1839 #label OSO
##cross-references EMBL:J04374
CLASSIFICATION #superfamily eggplant mosaic virus RNA-directed RNA
polymerase
KEYWORDS     ATP; nucleotidyltransferase; P-loop; RNA biosynthesis; RNA
replication
FEATURE      965-972   #region nucleotide-binding motif A (P-loop)\
1027-1032   #region nucleotide-binding motif B\
971         #binding_site ATP (Lys) #status Predicted
SUMMARY      #length 1839 #molecular-weight 204731 #checksum 7757

Query Match      64.0%; Score 55; DB 1; Length 1839;
Best Local Similarity 54.5%; Pred. No. 2.29e+00;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 160 SPALNTLYCSL 170
||||: |||
QY 1 SPALNMFQQL 11

RESULT 14
ENTRY   S52608 #type complete
TITLE   probable membrane protein YHL002c-a - yeast (Saccharomyces
cerevisiae)
ORGANISM #formal_name Saccharomyces cerevisiae
DATE     05-May-1995 #sequence_revision 19-Oct-1995 #text_change
ACCESSIONS S52608
REFERENCE   S46794
#authors   Favell, T.
#submission submitted to the EMBL Data Library, June 1994
#description The sequence of S. cerevisiae cosmid 9780.
#accession S52608
#molecule_type DNA
##residues 1-162 #label FAV
##cross-references EMBL:U10555; MRP:YHL002c-a
GENETICS
#map_position 8L
KEYWORDS   transmembrane protein
FEATURE    134-150
#domain transmembrane #status predicted #label TMM
SUMMARY    #length 162 #molecular-weight 17933 #checksum 1569

Query Match      62.8%; Score 54; DB 2; Length 162;
Best Local Similarity 71.4%; Pred. No. 3.60e+00;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 126 SKMYCOL 132
||||: |||
QY 5 NKMFQQL 11

RESULT 15
ENTRY   C71894 #type complete
TITLE   GTP cyclohydrolase ii/3,4-dihydroxy-2-butanone 4- phosphate
synthase - Helicobacter pylori (strain J99)
#formal_name Helicobacter pylori

```

```

#variety     strain J99
DATE         12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change
ACCESSIONS   C71894
REFERENCE     A71800
#authors     Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.;
Doig, P.C.; Smith, D.R.; Noonan, B.; Guild, B.C.; deJonge,
B.L.; Carmel, G.; Tummino, P.J.; Caruso, A.;
Uria-Nickelsen, M.; Mills, D.M.; Ives, C.; Gibson, R.;
Metzberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis,
G.F.; Trust, T.J.
#journal     Nature (1999) 397:176-180
#title       Genomic sequence comparison of two unrelated isolates of the
human gastric pathogen Helicobacter pylori.
#cross-references MUID:99120557
#accession   C71894
#status      preliminary
#molecule_type DNA
##residues  1-344 #label ARN
##cross-references GB:AE001505; GB:AE001439; NID:94155295; PID:94155296
##experimental_source strain J99
GENETICS
#gene        ribBA
CLASSIFICATION #superfamily riba bifunctional protein; 3,
4-dihydroxy-2-butanone 4-phosphate synthase homology;
cyclohydrolase homology
SUMMARY      #length 344 #molecular-weight 38890 #checksum 3146

Query Match      62.8%; Score 54; DB 2; Length 344;
Best Local Similarity 75.0%; Pred. No. 3.60e+00;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 205 SLKMFQQL 212
||||: |||
QY 3 ALNMFQQL 10

Search completed: Sat Apr 15 00:28:59 2000
Job time : 16 secs.

```

**THIS PAGE BLANK (USPTO)**

\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

mpsrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Sat Apr 15 00:29:16 2000; MasPar time 3.11 Seconds  
105.579 Million cell updates/sec  
Tabular output not generated.

```
>US-08-452-843-15
Description: (1-11) from US08452843.pep
Perfect score: 86
Sequence: 1 SPALNKMFCQL 11
```

Scoring table: PAM 150  
Gap 15

Searched: 82229 seqs, 29864866 residues

```
Post-processing: Minimum Match 0%
                  Listing first 45 summaries

Database:         swiss-prot38
                  1:swissprot
```

Statistics: Mean 25.100: Variance 28.852: scale 0.870

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query		Length	DB	ID	Description	Pred. No.
		Match						
1	86	100.0	207	1	P53_EQUAS		CELLULAR TUMOR ANTIGEN	1.31e-08
2	86	100.0	393	1	P53_HUMAN		CELLULAR TUMOR ANTIGEN	1.31e-08
3	82	95.3	393	1	P53_MACEA		CELLULAR TUMOR ANTIGEN	1.61e-07
4	82	95.3	393	1	P53_CERAE		CELLULAR TUMOR ANTIGEN	1.61e-07
5	82	95.3	393	1	P53_MACMU		CELLULAR TUMOR ANTIGEN	1.61e-07
6	77	89.5	280	1	P53_HORSE		CELLULAR TUMOR ANTIGEN	3.44e-06
7	77	89.5	314	1	P53_SPEBE		CELLULAR TUMOR ANTIGEN	3.44e-06
8	77	89.5	382	1	P53_SHEEP		CELLULAR TUMOR ANTIGEN	3.44e-06
9	77	89.5	386	1	P53_BOVIN		CELLULAR TUMOR ANTIGEN	3.44e-06
10	77	89.5	393	1	P53_CRICK		CELLULAR TUMOR ANTIGEN	3.44e-06
11	77	89.5	396	1	P53_WESAU		CELLULAR TUMOR ANTIGEN	3.44e-06
12	76	88.4	386	1	P53_FELCA		CELLULAR TUMOR ANTIGEN	6.29e-06
13	76	88.4	390	1	P53_MOUSE		CELLULAR TUMOR ANTIGEN	6.29e-06
14	75	87.2	363	1	P53_XENLA		CELLULAR TUMOR ANTIGEN	1.15e-05
15	75	87.2	373	1	P53_BRARE		CELLULAR TUMOR ANTIGEN	1.15e-05
16	75	87.2	396	1	P53_SALIR		CELLULAR TUMOR ANTIGEN	1.15e-05
17	72	83.7	381	1	P53_CANFA		CELLULAR TUMOR ANTIGEN	6.78e-05
18	71	82.6	391	1	P53_RABIT		CELLULAR TUMOR ANTIGEN	1.22e-04
19	65	75.6	391	1	P53_RAT		CELLULAR TUMOR ANTIGEN	3.74e-03
20	63	73.3	351	1	P53_ORYLA		CELLULAR TUMOR ANTIGEN	1.13e-02
21	63	73.3	367	1	P53_CHICK		CELLULAR TUMOR ANTIGEN	1.13e-02
22	55	64.0	366	1	P53_PLAPE		CELLULAR TUMOR ANTIGEN	7.16e-01
23	55	64.0	1839	1	POLR_EPVV		RNA REPLICASE POLYROT	7.16e-01

KW Nuclear protein; Phosphorylation; Apoptosis.  
 FT NON\_TER 1 1  
 FT DOMAIN 187 199 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 FT NON\_TER 207 207  
 SQ SEQUENCE 207 AA; 23428 MW; 0FBAE9C1 CRC32;  
 Query Match 100.0%; Score 86; DB 1; Length 207;  
 Best Local Similarity 100.0%; Pred. No. 1.31e-08;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 2 SPALNKMFCOL 12  
 |||||  
 QY 1 SPALNKMFCOL 11  
 RESULT 2  
 ID P53\_HUMAN STANDARD; PRT; 393 AA.  
 AC P04637;  
 DT 13-AUG-1987 (Rel. 05, Created)  
 DT 01-MAR-1989 (Rel. 10, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE CELLULAR TUMOR ANTIGEN P53 (PHOSPHOPROTEIN P53).  
 GN TP53.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;  
 OC Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 85230577.  
 RA ZAKUT-HOORI R., BIENZ-TADMOR B., GIVOL D., OREN M.;  
 RT "Human p53 cellular tumor antigen: cDNA sequence and expression in  
 RT COS cells.";  
 RL EMBO J. 4:1251-1255(1985).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 87064416.  
 RA LAMB P., CRAWFORD L.;  
 RT "Characterization of the human p53 gene.";  
 RL Mol. Cell. Biol. 6:1379-1385(1986).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 85267676.  
 RA HARLOW E., WILLIAMSON N.M., RALSTON R., HELFMAN D.M., ADAMS T.E.;  
 RT "Molecular cloning and in vitro expression of a cDNA clone for human  
 RT cellular tumor antigen p53.";  
 RL Mol. Cell. Biol. 5:1601-1610(1985).  
 RN [4]  
 RP TRANSFORMED HYBRIDOMA SV-80 CELL LINE, SEQUENCE FROM N.A.  
 RX MEDLINE; 87089826.  
 RA HARRIS N., BRILL E., SHOHAT O., PROKOCIMER M., WOLF D., ARAI N.,  
 RA ROTTNER V.;  
 RT "Molecular basis for heterogeneity of the human p53 protein.";  
 RL Mol. Cell. Biol. 6:4650-4656(1986).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 89108008.  
 RA BUCHANAN V.L., CHUMAKOV P.M., NINKINA N.N., SAMARINA O.P.,  
 RA GORGIEV G.P.;  
 RT "A variation in the structure of the protein-coding region of the  
 RT human p53 gene.";  
 RL Gene 70:245-252(1988).  
 RN [6]  
 RP SEQUENCE OF 101-393 FROM N.A.  
 RX MEDLINE; 85126934.  
 RA MATLASHESKI G., LAMB P., PIM D., PEACOCK J., CRAWFORD L.,  
 RA BENCHIMOL S.;  
 RT "Isolation and characterization of a human p53 cDNA clone: expression  
 RT of the human p53 gene.";  
 RL EMBO J. 3:3257-3262(1984).  
 RN [7]  
 RP NUCLEAR LOCALIZATION SIGNAL.  
 RX MEDLINE; 90191730.  
 RA ADDISON C., JENKINS J.R., STUR2BECHER H.-W.;

RT "The p53 nuclear localisation signal is structurally linked to a  
 RT p34cdc2 kinase motif.";  
 RL Oncogene 5:423-426(1990).  
 RN [8]  
 RP PHOSPHORYLATION BY P50/CDC2 AND CYCLIN B/CDC2.  
 RX MEDLINE; 90280456.  
 RA BISCHOFF J.R., FRIEDMAN P.N., MARSHAK D.R., PRIVES C., BEACH D.;  
 RT "Human p53 is phosphorylated by p50-cdc2 and cyclin B-cdc2.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 87:4766-4770(1990).  
 RN [9]  
 RP DEPHOSPHORYLATION BY PP2A.  
 RX MEDLINE; 91172186.  
 RA SCHEIDTMANN K.H., MUMBY M.C., RUNDELL K., WALTER G.;  
 RT "Dephosphorylation of simian virus 40 large-T antigen and p53 protein  
 RT by protein phosphatase 2A: inhibition by small-t antigen.";  
 RL Mol. Cell. Biol. 11:1996-2003(1991).  
 RN [10]  
 RP STRUCTURE BY NMR OF 319-360.  
 RX MEDLINE; 94294808.  
 RA CLORE G.M., OMICHINSKI J.G., SAKAGUCHI K., ZAMBRANO N., SAKAMOTO H.,  
 RA APPELLA E., GRONENBORN A.M.;  
 RT "High-resolution structure of the oligomerization domain of p53 by  
 RT multidimensional NMR.";  
 RL Science 265:386-391(1994).  
 RN [11]  
 RP STRUCTURE BY NMR OF 325-355.  
 RX MEDLINE; 95292092.  
 RA LEE W., HARVEY T.S., YIN Y., YAU P., LITCHFIELD D., ARROWSMITH C.H.;  
 RT "Solution structure of the tetrameric minimum transforming domain of  
 RT p53.";  
 RL Nat. Struct. Biol. 1:877-890(1994).  
 RN [12]  
 RP STRUCTURE BY NMR OF 326-354.  
 RX MEDLINE; 98026899.  
 RA MCCOY M., STAVRIDIS E.S., WATERMAN J.L., WIECZOREK A.M., OPELLA S.J.,  
 RA HALAZONETIS T.D.;  
 RT "Hydrophobic side-chain size is a determinant of the  
 RT three-dimensional structure of the p53 oligomerization domain.";  
 RL EMBO J. 16:6230-6236(1997).  
 RN [13]  
 RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 94-289.  
 RX MEDLINE; 94294806.  
 RA CHO Y., GORINA S., JEFFREY P.D., PAVLETICH N.P.;  
 RT "Crystal structure of a p53 tumor suppressor-DNA complex:  
 RT understanding tumorigenic mutations.";  
 RL Science 265:346-355(1994).  
 RN [14]  
 RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS) OF 13-29 IN COMPLEX WITH MDM2.  
 RX MEDLINE; 97081050.  
 RA KUSSIE P.H., GORINA S., MARECHAL V., ELENBAAS B., MOREAU J.,  
 RA LEVINE A.J., PAVLETICH N.P.;  
 RT "Structure of the MDM2 oncoprotein bound to the p53 tumor suppressor  
 RT transactivation domain.";  
 RL Science 274:948-953(1996).  
 RN [15]  
 RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 97-287 IN COMPLEX WITH 53BP2.  
 RX MEDLINE; 97035414.  
 RA GORINA S., PAVLETICH N.P.;  
 RT "Structure of the p53 tumor suppressor bound to the ankyrin and SH3  
 RT domains of 53BP2.";  
 RL Science 274:1001-1005(1996).  
 RN [16]  
 RP REVIEW.  
 RX MEDLINE; 94090335.  
 RA HARRIS C.C.;  
 RT "p53: at the crossroads of molecular carcinogenesis and risk  
 RT assessment.";  
 RL Science 262:1980-1981(1993).  
 RN [17]  
 RP REVIEW ON VARIANTS.  
 RX MEDLINE; 91289156.  
 RA HOOLSTEIN M., SIDRANSKY D., VOGELSTEIN B., HARRIS C.C.;  
 RT "p53 mutations in human cancers.";



Science 253:49-53(1991).  
[18]  
RN REVIEW ON VARIANTS.  
RX MEDLINE; 96271983.  
RA DE VRIES E.M.G., RIQUE D.O., DE VRIES T.N., HARTMANN A., BLASZYK H.,  
LIAO D., SOUSSI T., KOVACH J.S., SOMMER S.S.;  
RT "Database of mutations in the p53 and APC tumor suppressor genes  
designed to facilitate molecular epidemiological analyses.";  
RL Hum. Mutat. 7:202-213(1996).  
[19]  
RN VARIANT ARG-72.  
RX MEDLINE; 91153807.  
RA OLSCHWANG S., LAURENT-PUIG P., VASSAL A., SALMON R.-J., THOMAS G.;  
RT "Characterization of a frequent polymorphism in the coding sequence  
of the TP53 gene in colonic cancer patients and a control  
population.";  
RL Hum. Genet. 86:369-370(1991).  
[20]  
RN VARIANT LFS THR-133.  
RX MEDLINE; 92034774.  
RA LAW J.C., STRONG L.C., CHIDAMBARAM A., FERRELL R.E.;  
RT "A germ line mutation in exon 5 of the p53 gene in an extended cancer  
family";  
RL Cancer Res. 51:6385-6387(1991).  
[21]  
RN VARIANTS LFS CYS-245; TRP-248; PRO-252 AND LYS-258.  
RX MEDLINE; 91057657.  
RA MALKIN D., LI F.P., STRONG L.C., FRAUMENI J.F. JR., NELSON C.E.,  
KIM D.H., KASSEL J., GRKYA M.A., BISCHOFF F.Z., TAINSKY M.A.,  
FRIEND S.H.;  
RT "Germ line p53 mutations in a familial syndrome of breast cancer,  
sarcomas, and other neoplasms.";  
RL Science 250:1233-1238(1990).  
[22]  
RN VARIANT LFS ASP-245.  
RX MEDLINE; 91080929.  
RA SRIVASTAVA S., ZOU Z., PIROLLO K., BLATTNER W., CHANG E.H.;  
RT "Germ-line transmission of a mutated p53 gene in a cancer-prone  
family with Li-Fraumeni syndrome.";  
RL Nature 348:747-749(1990).  
[23]  
RN VARIANT LFS LEU-272.  
RX MEDLINE; 92147883.  
RA FELIX C.A., NAU M.M., TAKAHASHI T., MITSUDOMI T., CHIBA I.,  
POPLACK D.G., REAMAN G.H., COLE D.E., LETTERIO J.J., WHANG-PENG J.,  
KNUTSEN T., MINNA J.D.;  
RT "Hereditary and acquired p53 gene mutations in childhood acute  
lymphoblastic leukemia";  
RL J. Clin. Invest. 89:640-647(1992).  
[24]  
RN VARIANTS LFS HIS-273 AND VAL-325.  
RX MEDLINE; 92228023.  
RA MALKIN D., JOLLY K.W., BARRIER N., LOOK A.T., FRIEND S.H.,  
RA GEBHART M.C., ANDERSEN T.I., BORRESEN A.-L., LI F.P., GARBER J.,  
RA STRONG L.C.;  
RT "Germline mutations of the p53 tumor-suppressor gene in children and  
RT young adults with second malignant neoplasms";  
RL New Engl. J. Med. 326:1309-1315(1992).  
[25]  
RN VARIANTS BREAST TUMORS GLN-132; SER-249; LYS-280 AND LYS-285.  
RX MEDLINE; 90295284.  
RA BARTER J., IGGO R., GANNON J., LANE D.P.;  
RT "Genetic and immunochemical analysis of mutant p53 in human breast  
RT cancer cell lines";  
RL Oncogene 5:893-899(1990).  
[26]  
RN VARIANTS COLON TUMORS PHE-241 AND HIS-273.  
RX MEDLINE; 91017544.  
RA RODRIGUES N.R., ROWAN A., SMITH M.E.F., KERR I.B., BODMER W.F.,  
RA GANNON J.V., LANE D.P.;

Note: remainder of annotations omitted.

Query Match 100.0%; Score 86; DB 1; Length 393;  
Best Local Similarity 100.0%; Pred. No. 1.31e-08;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Db 127 SPALNMFQCL 137  
Qy 1 SPALNMFQCL 11  
RESULT 3  
ID P53 MACFA STANDARD; PRT; 393 AA.  
AC P56423;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 15-JUL-1998 (Rel. 36, Last annotation update)  
DE CELLULAR TUMOR ANTIGEN P53.  
GN TP53 OR P53.  
OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Primates; Catarrhini; Cercopitheidae; Cercopitheinae;  
OC Macaca.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA KHAN M.A., HANSEN C., WELSH J.A., BENNETT W.P.;  
RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES  
CC GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL  
CC CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN  
CC TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION. IT IS A  
CC TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION  
CC BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF  
CC THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.  
CC APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF  
CC BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2  
CC EXPRESSION.  
CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
CC -1- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY  
CC OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED  
CC IN MANY TYPES OF CANCER.  
CC -1- SIMILARITY: BELONGS TO THE P53 FAMILY.  
CC -----  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; U48957; AAB91535.1; -  
DR HSSP; P04637; 1SAH.  
DR PROSITE; PS00348; P53; 1.  
DR PFAM; PF00870; P53; 1.  
KW Anti-oncogene; DNA-binding; Transcription regulation; Activator;  
KW Nuclear protein; Phosphorylation; Apoptosis.  
FT DOMAIN 1 80 ASP/GLU-RICH (ACIDIC).  
FT DOMAIN 81 150 HYDROPHOBIC.  
FT DOMAIN 319 393 HIGHLY BASIC AND MAY BE INVOLVED IN  
FT INTERACTION WITH DNA.  
FT DOMAIN 311 323 NUCLEAR LOCALIZATION SIGNAL.  
FT MOD\_RES 392 392 PHOSPHORYLATION (BY SIMILARITY).  
SQ SEQUENCE 393 AA; 43678 MW; 2499AC47 CRC32;  
Query Match 95.3%; Score 82; DB 1; Length 393;  
Best Local Similarity 90.9%; Pred. No. 1.61e-07;  
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Db 127 SPALNMFQCL 137  
Qy 1 SPALNMFQCL 11  
RESULT 4

ID P53\_CERAE STANDARD; PRT; 393 AA.  
AC P13481;  
DT 01-JAN-1990 (Rel. 13, Created)  
DT 01-JAN-1990 (Rel. 13, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE CELLULAR TUMOR ANTIGEN P53.  
GN TP53.  
OS Cercopithecus aethiops (Green monkey) (Grivet).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Primates; Catarrhini; Cercopithecidae; Cercopithecinae;  
OC Chlorocebus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA TISSUE=LIVER;  
RX MEDLINE; 90045967.  
RT RIGAUDY P., ECKHARDT W.;  
RT "Nucleotide sequence of a cDNA encoding the monkey cellular  
RT phosphoprotein p53.";  
RL Nucleic Acids Res. 17:8375-8375(1989).  
CC -1- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES  
CC GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL  
CC CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN  
CC TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION, IT IS A  
CC TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION  
CC BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF  
CC THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.  
CC APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF  
CC BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2  
CC EXPRESSION.  
CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
CC -1- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY  
CC OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED  
CC IN MANY TYPES OF CANCER.  
CC -1- SIMILARITY: BELONGS TO THE P53 FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; X16384; CAA34420.1; -  
DR PIR; S06594; S06594.  
DR HSSP; P04637; 1SAH.  
DR PROSITE; PS00348; P53; 1.  
DR PFAM; PF00870; P53; 1.  
KW Anti-oncogene; DNA-binding; Transcription regulation; Activator;  
FT DOMAIN 1 68 ASP/GLU-RICH (ACIDIC).  
FT DOMAIN 81 150 HYDROPHOBIC.  
FT DOMAIN 319 393 HIGHLY BASIC AND MAY BE INVOLVED IN  
FT INTERACTION WITH DNA.  
FT DOMAIN 311 323 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
FT MOD\_RES 392 392 PHOSPHORYLATION (BY SIMILARITY).  
SQ SEQUENCE 393 AA; 43696 MW; BBE7DC62 CRC32;  
Query Match 95.38; Score 82; DB 1; Length 393;  
Best Local Similarity 90.9%; Pred. No. 1.61e-07;  
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Db 127 SPDLNKMFCOL 137  
QY 1 SPALNKMFCOL 11  
RESULT 5  
ID P53\_MACMU STANDARD; PRT; 393 AA.  
AC P56424;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)

DE CELLULAR TUMOR ANTIGEN P53.  
GN TP53 OR P53.  
OS Macaca mulatta (Rhesus macaque).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Primates; Catarrhini; Cercopithecidae; Cercopithecinae;  
OC Macaca.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA KHAN M.A., HANSEN C., WELSH J.A., BENNETT W.P.;  
RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES  
CC GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL  
CC CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN  
CC TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION, IT IS A  
CC TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION  
CC BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF  
CC THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.  
CC APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF  
CC BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2  
CC EXPRESSION.  
CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
CC -1- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY  
CC OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED  
CC IN MANY TYPES OF CANCER.  
CC -1- SIMILARITY: BELONGS TO THE P53 FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; U48956; AAB91534.1; -  
DR HSSP; P04637; 1SAH.  
DR PROSITE; PS00348; P53; 1.  
DR PFAM; PF00870; P53; 1.  
KW Anti-oncogene; DNA-binding; Transcription regulation; Activator;  
FT DOMAIN 1 80 ASP/GLU-RICH (ACIDIC).  
FT DOMAIN 81 150 HYDROPHOBIC.  
FT DOMAIN 319 393 HIGHLY BASIC AND MAY BE INVOLVED IN  
FT INTERACTION WITH DNA.  
FT DOMAIN 311 323 NUCLEAR LOCALIZATION SIGNAL.  
FT MOD\_RES 392 392 PHOSPHORYLATION (BY SIMILARITY).  
SQ SEQUENCE 393 AA; 43655 MW; 11A9B7F8 CRC32;  
Query Match 95.38; Score 82; DB 1; Length 393;  
Best Local Similarity 90.9%; Pred. No. 1.61e-07;  
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Db 127 SPDLNKMFCOL 137  
QY 1 SPALNKMFCOL 11  
RESULT 6  
ID P53\_HORSE STANDARD; PRT; 280 AA.  
AC P79892; Q29481;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).  
GN TP53 OR P53.  
OS Equus caballus (Horse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Perissodactyla; Equidae; Equus.  
RN [1]  
RP SEQUENCE OF 1-263 FROM N.A.  
RC TISSUE=SPLEEN;  
RX MEDLINE; 97070350.  
RA FAZZI K.A., KRAEGL S.A., GRIFFEY S.M., THEON A.P., MADEWELL B.R.;



CC TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION, IT IS A  
 CC TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION  
 CC BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF  
 CC THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.  
 CC APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF  
 CC BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2  
 CC EXPRESSION.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -1- SIMILARITY: BELONGS TO THE P53 FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; X81705; CAA57349.1; -  
 CC HSSP; P04637; 1PPT.  
 CC PROSITE; PS00348; P53; 1.  
 CC PFAM; PF00870; P53; 1.  
 CC Anti-oncogene; DNA-binding; Transcription regulation; Activator;  
 CC Nuclear protein; Phosphorylation; Apoptosis.  
 CC DOMAIN 1 66 ASP/GLU-RICH (ACIDIC).  
 CC FT MOD\_RES 300 312 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 CC FT MOD\_RES 381 381 PHOSPHORYLATION (BY SIMILARITY).  
 CC SEQUENCE 382 AA; 42809 MW; 0CB99A00 CRC32;  
 CC -----  
 CC Query Match 89.5%; Score 77; DB 1; Length 382;  
 CC Best Local Similarity 81.8%; Pred. No. 3.44e-06;  
 CC Matches 9; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
 CC -----  
 CC Db 115 SPLNKLFCQL 125  
 CC ||:|||||  
 CC QY 1 SPALNKMFCQL 11  
 CC -----  
 CC RESULT 9  
 CC ID P53\_BOVIN STANDARD; PRT; 386 AA.  
 CC AC Q2628;  
 CC DT 01-NOV-1997 (Rel. 35, Created)  
 CC DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 CC DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 CC DE CELLULAR TUMOR ANTIGEN P53.  
 CC TP53.  
 CC OS Bos taurus (Bovine), and Bos indicus (Zebu).  
 CC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 CC OC Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae;  
 CC OC Bos.  
 CC RN [1]  
 CC RP SEQUENCE FROM N.A.  
 CC RC SPECIES-BOVINE; TISSUE=LIVER;  
 CC RX MEDLINE; 95352829.  
 CC RA DEQUIEDT F., KETTMANN R., BURNY A., WILLEMS L.;  
 CC RT "Nucleotide sequence of the bovine P53 tumor-suppressor cDNA.";  
 CC RL DNA Seq. 5:261-264(1995).  
 CC RN [2]  
 CC RP SEQUENCE OF 13-386 FROM N.A.  
 CC RC SPECIES-BOVINE; STRAIN=HOLSTEIN; TISSUE=THYMUS;  
 CC RX MEDLINE; 96401400.  
 CC RA KOMORI H., ISHIGURO N., HORIUCHI M., SHINAGAWA M., AIDA Y.;  
 CC RT "Predominant p53 mutations in enzootic bovine leukemic cell lines.";  
 CC RL Vet. Immunol. Immunopathol. 52:53-63(1996).  
 CC RN [3]  
 CC RP SEQUENCE FROM N.A.  
 CC RC SPECIES-B.INDICUS; STRAIN=BORAN; TISSUE=BLOOD;  
 CC RA BISHOP R.R.P., GOBRIGHT E.E.I.;  
 CC RL Submitted (Apr-1997) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES  
 CC GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL  
 CC CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN  
 CC TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION, IT IS A

CC TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION  
 CC BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF  
 CC THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.  
 CC APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF  
 CC BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2  
 CC EXPRESSION.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -1- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY  
 CC OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED  
 CC IN MANY TYPES OF CANCER.  
 CC -1- SIMILARITY: BELONGS TO THE P53 FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; X81704; CAA57348.1; -  
 CC EMBL; D49825; BAA08629.1; -  
 CC EMBL; U74486; AAB51214.1; -  
 CC HSSP; P04637; 1YCR.  
 CC PROSITE; PS00348; P53; 1.  
 CC PFAM; PF00870; P53; 1.  
 CC Anti-oncogene; DNA-binding; Transcription regulation; Activator;  
 CC Nuclear protein; Phosphorylation; Apoptosis.  
 CC DOMAIN 1 59 ASP/GLU-RICH (ACIDIC).  
 CC FT MOD\_RES 304 316 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 CC FT MOD\_RES 385 385 PHOSPHORYLATION (BY SIMILARITY).  
 CC FT CONFLICT 380 380 R -> T (IN REF. 2).  
 CC SEQUENCE 386 AA; 43255 MW; 0322BF3D CRC32;  
 CC -----  
 CC Query Match 89.5%; Score 77; DB 1; Length 386;  
 CC Best Local Similarity 81.8%; Pred. No. 3.44e-06;  
 CC Matches 9; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
 CC -----  
 CC Db 119 SPLNKLFCQL 129  
 CC ||:|||||  
 CC QY 1 SPALNKMFCQL 11  
 CC -----  
 CC RESULT 10  
 CC ID P53\_CRIGR STANDARD; PRT; 393 AA.  
 CC AC 009185; O64397; P97258; P97788;  
 CC DT 01-NOV-1997 (Rel. 35, Created)  
 CC DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 CC DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 CC DE CELLULAR TUMOR ANTIGEN P53.  
 CC GN TP53 OR P53.  
 CC OS Cricetulus griseus (Chinese hamster).  
 CC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 CC OC Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae; Cricetulus.  
 CC RN [1]  
 CC RP SEQUENCE FROM N.A.  
 CC RA CHAUNG W., MI L.J., BOORSTEIN R.J.;  
 CC RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.  
 CC RN [2]  
 CC RP SEQUENCE FROM N.A.  
 CC RC TISSUE=LIVER;  
 CC RX MEDLINE; 97183659.  
 CC RA LEE H., LARNER J.M., HAMLIN J.L.;  
 CC RT "Cloning and characterization of Chinese hamster p53 cDNA.";  
 CC RL Gene 184:177-183(1997).  
 CC RN [3]  
 CC RP SEQUENCE FROM N.A.  
 CC RC TISSUE=EMBRYONIC FIBROBLAST;  
 CC RA SHIMIZU T., NIKAIKO O., SUZUKI F.;  
 CC RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES  
 CC GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL  
 CC CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN

CC TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION, IT IS A  
CC TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION  
CC BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF  
CC THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.  
CC APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF  
CC BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2  
CC EXPRESSION.  
CC SUBCELLULAR LOCATION: NUCLEAR.  
CC -1- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY  
CC OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED  
CC IN MANY TYPES OF CANCER.  
CC -1- SIMILARITY: BELONGS TO THE P53 FAMILY.  
CC  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: Y08900; CAA70108.1; -  
CC EMBL: Y08901; CAA70109.1; -  
CC EMBL: U50395; AAC53040.1; -  
CC EMBL: D86070; BAAL3004.1; -  
CC HSSP: P04637; 1YCO.  
CC PROSITE: PS00348; P53; 1.  
CC PFAM: PF00870; P53; 1.  
CC Anti-oncogene; DNA-binding; Transcription regulation; Activator;  
CC Nuclear protein; Phosphorylation; Apoptosis.  
CC DOMAIN 1 74 ASP/GLU-RICH (ACIDIC).  
CC FT DOMAIN 75 150 HYDROPHOBIC.  
CC FT DOMAIN 316 390 HIGHLY BASIC AND MAY BE INVOLVED IN  
CC FT DOMAIN 311 323 INTERACTION WITH DNA.  
CC FT DOMAIN 323 392 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
CC MOD\_RES 392 392 PHOSPHORYLATION (BY SIMILARITY).  
CC VARIANT 133 133 L -> Q (IN CELL LINE V79-4).  
CC VARIANT 135 135 C -> W (IN CELL LINE V79-4).  
CC CONFLICT 103 103 Y -> F (IN REF. 2).  
CC SEQUENCE 393 AA; 43378 MW; 402EB149 CRC32;  
Query Match 89.58; Score 77; DB 1; Length 393;  
Best Local Similarity 81.88; Pred. No. 3.44e-06;  
Matches 9; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
Db 127 SPSLNKLFCQL 137  
||:|||||  
QY 1 SPALNKMFCQL 11  
RESULT 11  
ID P53\_MESAU STANDARD; PRT; 396 AA.  
AC Q00366; P97276;  
DT 01-DEC-1992 (Rel. 24, Created)  
DT 01-DEC-1992 (Rel. 24, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE CELLULAR TUMOR ANTIGEN P53.  
GN TP53.  
OS Mesocricetus auratus (Golden hamster).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae; Mesocricetus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=SYRIAN; TISSUE=KIDNEY;  
RX MEDLINE: 92210007.  
RA LEGROS Y., MCINTYRE P., SOUSSI T.;  
RT "The cDNA cloning and immunological characterization of hamster p53.";  
RL Gene 112:247-250(1992).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA HOU E.W., WISEMAN R.;  
RL Submitted (APR-1994) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES

CC GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL  
CC CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN  
CC TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION, IT IS A  
CC TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION  
CC BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF  
CC THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.  
CC APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF  
CC BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2  
CC EXPRESSION.  
CC SUBCELLULAR LOCATION: NUCLEAR.  
CC -1- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY  
CC OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED  
CC IN MANY TYPES OF CANCER.  
CC -1- SIMILARITY: BELONGS TO THE P53 FAMILY.  
CC  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: M75144; AAA37085.1; -  
CC EMBL: U07182; AAB41344.1; -  
CC PIR: JH0633; JH0633.  
CC HSSP: P04637; 1YCO.  
CC PROSITE: PS00348; P53; 1.  
CC PFAM: PF00870; P53; 1.  
CC Anti-oncogene; DNA-binding; Transcription regulation; Activator;  
CC Nuclear protein; Phosphorylation; Apoptosis.  
CC DOMAIN 1 77 ASP/GLU-RICH (ACIDIC).  
CC FT DOMAIN 78 153 HYDROPHOBIC.  
CC FT DOMAIN 319 393 HIGHLY BASIC AND MAY BE INVOLVED IN  
CC FT DOMAIN 314 326 INTERACTION WITH DNA.  
CC MOD\_RES 395 395 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
CC CONFLICT 188 188 G -> S (IN REF. 2).  
CC SEQUENCE 396 AA; 43631 MW; C2668ADE CRC32;  
Query Match 89.58; Score 77; DB 1; Length 396;  
Best Local Similarity 81.88; Pred. No. 3.44e-06;  
Matches 9; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
Db 130 SPSLNKLFCQL 140  
||:|||||  
QY 1 SPALNKMFCQL 11  
RESULT 12  
ID P53\_FELCA STANDARD; PRT; 386 AA.  
AC P41685;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE CELLULAR TUMOR ANTIGEN P53.  
GN TP53.  
OS Felis silvestris catus (Cat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Carnivora; Fissipedia; Felidae; Felis.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=LYMPH NODE;  
RX MEDLINE: 94333960.  
RA OKUDA M., UEDA A., SAKAI T., OHASHI T., MOMOI Y., YOUN H.Y.,  
RA WATARI T., GOITSUKA R., TSUJIMOTO H., HASEGAWA A.;  
RT "Cloning of feline p53 tumor-suppressor gene and its aberration in  
RL hematopoietic tumors.";  
RL Int. J. Cancer 58:602-607(1994).  
RN [2]  
RP SEQUENCE OF 34-354 FROM N.A.  
RX MEDLINE: 94114699.  
RA OKUDA M., UEDA A., MATSUMOTO Y., MOMOI Y., WATARI T., GOITSUKA R.,

RA O'BRIEN S.J., TSUJIMOTO H., HASEGAWA A.;  
RT "Molecular cloning and chromosomal mapping of feline p53 tumor  
RL suppressor gene";  
RL J. Vet. Med. Sci. 55:801-805(1993).  
CC -1- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES  
CC GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL  
CC CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN  
CC TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION, IT IS A  
CC TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION  
CC BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF  
CC THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.  
CC APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF  
CC BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2  
CC EXPRESSION.  
CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
CC -1- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY  
CC OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED  
CC IN MANY TYPES OF CANCER.  
CC -1- SIMILARITY: BELONGS TO THE P53 FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; D26608; BAA05653.1; -  
CC EMBL; D16460; BAA03927.1; -  
CC HSSP; P04637; ISAH.  
CC PROSITE; PS00348; P53; 1.  
CC PAM; PF00870; P53; 1.  
CC Anti-oncogene; DNA-binding; Transcription regulation; Activator;  
CC Nuclear protein; Phosphorylation; Apoptosis.  
CC DOMAIN 1 59 ASP/GLU-RICH (ACIDIC).  
CC FT DOMAIN 304 316 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
CC FT MOD\_RES 385 385 PHOSPHORYLATION (BY SIMILARITY).  
CC FT CONFLICT 285 285 K -> R (IN REF. 2).  
CC SEQUENCE 386 AA; 42692 MW; D6C7132A CRC32;  
Query Match 88.4%; Score 76; DB 1; Length 386;  
Best Local Similarity 81.8%; Pred. No. 6.29e-06;  
Matches 9; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
Db 119 SPLNKLFCQL 129  
QY 1 SPALNMFQQL 11  
RESULT 13  
ID P53 MOUSE STANDARD; PRT; 390 AA.  
AC P02340;  
DT 21-JUL-1986 (Rel. 01, Created)  
DI 01-NOV-1990 (Rel. 16, Last sequence update)  
DE 01-NOV-1997 (Rel. 35, Last annotation update)  
DE CELLULAR TUMOR ANTIGEN P53.  
GN TP53 OR TRP53 OR P53.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
CC [1]  
CC SEQUENCE FROM N.A.  
CC MEDLINE; 85027173.  
RA BIENZ B., ZAKUT-HOURI R., GIVOL D., OREN M.;  
RT "Analysis of the gene coding for the murine cellular tumour antigen  
RT p53";  
RL EMBL J. 3:2179-2183(1984).  
CC [2]  
CC SEQUENCE FROM N.A.  
CC MEDLINE; 8408204.  
RA ZAKUT-HOURI R., OREN M., BIENZ B., LAVIE V., HAZUM S., GIVOL D.;  
RT "A single gene and a pseudogene for the cellular tumour antigen p53";  
-----  
RL Nature 306:594-597(1983).  
RN [3]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 84272240.  
RA JENKINS J.R., RUDGE K., REDMOND S., WADE-EVANS A.;  
RT "Cloning and expression analysis of full length mouse cDNA sequences  
RT encoding the transformation associated protein p53";  
RL Nucleic Acids Res. 12:5609-5626(1984).  
CC [4]  
CC SEQUENCE FROM N.A. (CLONES PCD53; P53-M11 AND P53-M8).  
RX MEDLINE; 87064640.  
RA ARAI N., NOMURA D., YOKOTA K., WOLF D., BRILL E., SHOHAT O.,  
RA ROTTER V.;  
RT "Immunologically distinct p53 molecules generated by alternative  
RT splicing";  
RL Mol. Cell. Biol. 6:3232-3239(1986).  
RN [5]  
RP SEQUENCE OF 222-258 FROM N.A.  
RX MEDLINE; 92115342.  
RA BURNS P.A., KEMP C.J., GANNON J.V., LANE D.P., BREMMER R.;  
RT "Loss of heterozygosity and mutational alterations of the p53 gene in  
RT skin tumours of interspecific hybrid mice";  
RL Oncogene 6:2363-2369(1991).  
RN [6]  
RP PHOSPHORYLATION SITES.  
RX MEDLINE; 86149247.  
RA SAMAD A., ANDERSON C.W., CARROLL R.B.;  
RT "Mapping of phosphonoester and apparent phosphodiester bonds of the  
RT oncogene product p53 from simian virus 40-transformed 3T3 cells";  
RL Proc. Natl. Acad. Sci. U.S.A. 83:897-901(1986).  
RN [7]  
RP PHOSPHORYLATION SITES.  
RX MEDLINE; 91006019.  
RA MEER D.W., SIMON S., KIKKAWA U., ECKHART W.;  
RT "The p53 tumour suppressor protein is phosphorylated at serine 389 by  
RT casein kinase II";  
RL EMBL J. 9:3253-3260(1990).  
CC -1- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES  
CC GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL  
CC CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN  
CC TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION, IT IS A  
CC TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION  
CC BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF  
CC THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.  
CC APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF  
CC BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2  
CC EXPRESSION.  
CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
CC -1- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY  
CC OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED  
CC IN MANY TYPES OF CANCER.  
CC -1- SIMILARITY: BELONGS TO THE P53 FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; X00876; CAA25420.1; -  
CC EMBL; X00877; CAA25420.1; JOINED.  
CC EMBL; X00878; CAA25420.1; JOINED.  
CC EMBL; X00879; CAA25420.1; JOINED.  
CC EMBL; X00880; CAA25420.1; JOINED.  
CC EMBL; X00881; CAA25420.1; JOINED.  
CC EMBL; X00882; CAA25420.1; JOINED.  
CC EMBL; X00883; CAA25420.1; JOINED.  
CC EMBL; X00884; CAA25420.1; JOINED.  
CC EMBL; X00885; CAA25420.1; JOINED.  
CC EMBL; K01700; AAA39884.1; -

```
DR EMBL; X01237; CAA25625.1; -
DR EMBL; X00741; CAA25323.1; -
DR EMBL; M13872; AAA39881.1; -
DR EMBL; M13873; AAA39882.1; -
DR EMBL; M13874; AAA39883.1; ALT_SEQ.
DR EMBL; S77930; AAB21108.1; -
DR PIR; A02684; DNMS53.
DR PIR; A22739; A22739.
DR PIR; S38822; S38822.
DR HSP; P04637; IPET.
DR TRANSFAC; T01806; -.
DR MGD; MGI-98834; TRP53.
DR PROSITE; PS00348; P53; 1.
DR PFAM; PF00870; P53; 1.
KW Anti-oncogene; DNA-binding; Transcription regulation; Activator;
KW Nuclear protein; Phosphorylation; Apoptosis; Disease mutation.
FT DOMAIN 1 75 ASP/GLU-RICH (ACIDIC).
FT DOMAIN 76 150 HYDROPHOBIC.
FT DOMAIN 276 390 HIGHLY BASIC AND MAY BE INVOLVED IN
INTERACTION WITH DNA.
FT DOMAIN 308 320 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT MOD_RES 312 312 PHOSPHORYLATION.
FT MOD_RES 389 389 PHOSPHORYLATION (BY CK2).
FT VARIANT 135 135 A->V (CAN COOPERATE WITH AN ACTIVATED
RAS TO TRANSFORM FIBROBLASTS).
FT VARIANT 168 168 E->G (IN CLONE P53-M11).
FT CONFLICT 48 48 Q->R (IN REF. 3).
FT CONFLICT 79 81 PVA->QW (IN REF. 3).
SQ SEQUENCE 390 AA; 43458 MW; 8943DD93 CRC32;

Query Match 88.4%; Score 76; DB 1; Length 390;
Best Local Similarity 81.8%; Pred. No. 6.29e-06;
Matches 9; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 124 SPPLNKLFCOL 134
QY 1 SPALNKMFCOL 11

RESULT 14
ID P53_XENLA STANDARD; PRT; 363 AA.
AC P07193;
DT 01-APR-1988 (Rel. 07, Created)
DT 01-APR-1988 (Rel. 07, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE CELLULAR TUMOR ANTIGEN P53.
GN TP53.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Amphibia;
OC Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae; Xenopodinae;
OC Xenopus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 88143684.
RA SOUSSI I., DE FROMENTEL C.C., MECHALI M., MAY P., KRESS M.;
RT "Cloning and characterization of a cDNA from Xenopus laevis coding
for a protein homologous to human and murine p53.";
RL Oncogene 1:71-78(1987).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 94134403.
RA HOEVER M., CLEMENT J.H., WEDLICH D., MONTENARH M., KNOCHEL W.;
RT "Overexpression of wild-type p53 interferes with normal development
in Xenopus laevis embryos.";
RL Oncogene 9:109-120(1994).
CC -1- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES
GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL
CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN
TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION, IT IS A
TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION
BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF
THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.
APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF
BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2
EXPRESSION (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- SIMILARITY: BELONGS TO THE P53 FAMILY.

CC BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2
EXPRESSION (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- TISSUE SPECIFICITY: UBIQUITOUS.
CC -1- SIMILARITY: BELONGS TO THE P53 FAMILY.
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC EMBL; M36962; AAA49923.1; -
CC EMBL; X05191; CAA28821.1; -
CC EMBL; X77546; CAA54672.1; -
CC EMBL; S68353; AAC50746.1; -
CC PIR; A29376; A29376.
CC HSP; P04637; IPTSR.
CC PROSITE; PS00348; P53; 1.
CC PFAM; PF00870; P53; 1.
KW Anti-oncogene; DNA-binding; Transcription regulation; Activator;
KW Nuclear protein; Phosphorylation; Apoptosis.
FT DOMAIN 281 293 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT MOD_RES 362 362 T->S (IN REF. 2).
FT CONFLICT 52 52 T->S (IN REF. 2).
FT CONFLICT 71 71 MISSING (IN REF. 2).
FT CONFLICT 296 296 MISSING (IN REF. 2).
SQ SEQUENCE 363 AA; 40692 MW; 75D7D796 CRC32;

Query Match 87.2%; Score 75; DB 1; Length 363;
Best Local Similarity 81.8%; Pred. No. 1.15e-05;
Matches 9; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 101 SPPLNKLFCOL 111
QY 1 SPALNKMFCOL 11

RESULT 15
ID P53_BRARE STANDARD; PRT; 373 AA.
AC P79734;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE CELLULAR TUMOR ANTIGEN P53.
GN TP53.
OS Brachydanio rerio (Zebrafish) (Zebra danio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Actinopterygii;
OC Neopterygii; Teleostei; Euteleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Cyprininae; Rasbora; Rasbora; Danio.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 97344388.
RA CHONG R., FORD B.L., O'NEAL P.E., MATHEWS C.Z., BRADFORD C.S.,
RA THONGTAN T., BARNES D.W., HENDRICKS J.D., BAILEY G.S.;
RT "Zebrafish (Danio rerio) p53 tumor suppressor gene: cDNA sequence and
expression during embryogenesis.";
RL Mol. Mar. Biol. Biotechnol. 6:88-97(1997).
CC -1- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES
GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL
CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN
TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION, IT IS A
TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION
BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF
THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.
APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF
BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2
EXPRESSION (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- SIMILARITY: BELONGS TO THE P53 FAMILY.
```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; U60804; AAB40617.1; -  
DR HSP; P04637; LISR  
DR ZFIN; ZDB-GENE-990415-32; TP53..  
DR PROSITE; PS00348; P53; 1.  
DR PFAM; PF00870; P53; 1.  
KW Anti-oncogene; DNA-binding; Transcription regulation; Activator;  
KW Nuclear protein; Phosphorylation; Apoptosis.  
FT DOMAIN 280 296 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
FT MOD\_RES 372 372 PHOSPHORYLATION (BY SIMILARITY).  
SQ SEQUENCE 373 AA; 41899 MW; 706A4B9C CRC32;


Query Match 87.2%; Score 75; DB 1; Length 373;  
Best Local Similarity 81.8%; Pred. No. 1.15e-05;  
Matches 9; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 95 SPDLNKFQQL 105  
QY 1 SPALNMFQQL 11

Search completed: Sat Apr 15 00:29:59 2000  
Job time : 43 secs.



\*\*\*\*\*



(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.  
 Copyright (c) 1993-1998 University of Edinburgh, U.K.  
 Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Sat Apr 15 00:30:16 2000; MasPar time 7.28 Seconds  
 104.771 Million cell updates/sec  
 Tabular output not generated.

Title: >US-08-452-843-15  
 Description: (1-11) from US08452843.pep  
 Perfect Score: 86  
 Sequence: 1 SPALNKMFCQL 11

Scoring table: PAM 150  
 Gap 15

Searched: 225878 seqs, 69334122 residues

Post-processing: Minimum Match 0%  
 Listing first 45 summaries

Database: sptrembl12

1:sp-archaea 2:sp-bacteria 3:sp-fungi 4:sp-human  
 5:sp-invertebrate 6:sp-mammal 7:sp-mhc 8:sp-organelle  
 9:sp-phage 10:sp-plant 11:sp-rodent 12:sp\_unclassified  
 13:sp-vertebrate 14:sp-virus

Statistics: Mean 24.303; Variance 29.401; scale 0.827

Pred. No. is the number of results predicted by chance to have a  
 score greater than or equal to the score of the result being printed,  
 and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	86	100.0	393	4	Q15087 P53 TRANSFORMATION SUP	5.79e-08
2	86	100.0	393	4	Q15088 P53 TRANSFORMATION SUP	5.79e-08
3	86	100.0	393	4	Q15086 P53 TRANSFORMATION SUP	5.79e-08
4	86	100.0	393	4	Q16810 CELLULAR TUMOR ANTIGEN	5.79e-08
5	86	100.0	393	4	Q16807 CELLULAR TUMOR ANTIGEN	5.79e-08
6	86	100.0	393	4	Q16808 CELLULAR TUMOR ANTIGEN	5.79e-08
7	86	100.0	393	4	Q16835 P53 TRANSFORMATION SUP	5.79e-08
8	86	100.0	393	4	Q16809 CELLULAR TUMOR ANTIGEN	5.79e-08
9	86	100.0	393	4	Q16848 CELLULAR TUMOR ANTIGEN	5.79e-08
10	86	100.0	393	4	Q16811 CELLULAR TUMOR ANTIGEN	5.79e-08
11	77	89.5	205	11	Q35873 CELLULAR TUMOR ANTIGEN	1.25e-05
12	77	89.5	238	14	P89004 P53 (FRAGMENT)	1.25e-05
13	77	89.5	286	14	P90332 P53 (FRAGMENT)	1.25e-05
14	77	89.5	286	14	P89003 P53 (FRAGMENT)	1.25e-05
15	77	89.5	378	14	P89002 P53 (FRAGMENT)	1.25e-05
16	77	89.5	391	6	Q36006 CELLULAR TUMOR ANTIGEN	1.25e-05
17	76	88.4	390	11	O70366 CELLULAR TUMOR ANTIGEN	2.23e-05
18	76	88.4	391	11	O9WUR6 CELLULAR TUMOR ANTIGEN	2.23e-05
19	75	87.2	265	13	Q9W681 CELLULAR TUMOR ANTIGEN	3.98e-05
20	75	87.2	265	13	Q9W680 CELLULAR TUMOR ANTIGEN	3.98e-05

21	75	87.2	265	13	Q9W682	CELLULAR TUMOR ANTIGEN	3.98e-05
22	75	87.2	376	13	Q93379	CELLULAR TUMOR ANTIGEN	3.98e-05
23	72	83.7	281	6	Q29475	CELLULAR TUMOR ANTIGEN	2.22e-04
24	72	83.7	285	6	Q95326	CELLULAR TUMOR ANTIGEN	2.22e-04
25	68	79.1	369	13	Q9W678	CELLULAR TUMOR ANTIGEN	2.09e-03
26	66	76.7	135	11	Q64451	CELLULAR TUMOR ANTIGEN	6.25e-03
27	61	70.9	1520	5	O15829	CARBAMYL PHOSPHATE SYN	9.04e-02
28	60	69.8	367	13	Q9W679	CELLULAR TUMOR ANTIGEN	1.52e-01
29	60	69.8	430	5	O18301	C29E6.3 PROTEIN.	1.52e-01
30	60	69.8	466	5	Q23602	ZK809.1 PROTEIN.	1.52e-01
31	58	67.4	342	13	O57538	CELLULAR TUMOR ANTIGEN	4.24e-01
32	58	67.4	342	13	Q92143	CELLULAR TUMOR ANTIGEN	4.24e-01
33	58	67.4	497	11	Q9W0J0	P73 (FRAGMENT)	4.24e-01
34	58	67.4	499	4	O15351	P73 PROTEIN.	4.24e-01
35	58	67.4	636	4	O15350	P53-LIKE TRANSCRIPTION	4.24e-01
36	58	67.4	637	6	Q9XSK8	P53-LIKE TRANSCRIPTION	4.24e-01
37	58	67.4	641	13	Q9W664	P73.	4.24e-01
38	55	64.0	327	14	Q9W6Y3	POLYPROTEIN (FRAGMENT)	1.90e+00
39	55	64.0	360	5	Q20723	F53F4.7 PROTEIN.	1.90e+00
40	54	62.8	344	2	O25484	GTP CYCLOHYDROLASE II/	3.09e+00
41	54	62.8	344	2	Q92140	GTP CYCLOHYDROLASE II/	3.09e+00
42	54	62.8	389	13	Q9W623	NEK2B.	3.09e+00
43	54	62.8	442	13	Q9W622	NEK2A.	3.09e+00
44	54	62.8	587	5	Q21432	COSMID K11G12.	3.09e+00
45	53	61.6	680	11	O88898	TA*P53 ALPHA.	5.01e+00

## ALIGNMENTS

RESULT 1  
 ID Q15087 PRELIMINARY; PRT; 393 AA.  
 AC Q15087;  
 DT 01-NOV-1996 (TREMBlrel. 01, Created)  
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
 DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)  
 DE P53 TRANSFORMATION SUPPRESSOR (FRAGMENT).  
 GN P53.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 92007731.  
 RA FARRELL P.J., ALLAN G., SHANAHAN F., VOUSDEN K.H., CROOK T.;  
 RT "p53 is frequently mutated in Burkitt's lymphoma cell lines.";  
 RL EMBO J. 10:2879-2887(1991).  
 DR EMBL; X60014; CAA42629.1; -  
 DR HSSP; F04637; ISAH.  
 DR PFAM; PF00870; P53; 1.  
 FT VARIANT 237 237 I -> M.  
 FT NON\_TER 393 393  
 SQ SEQUENCE 393 AA; 43694 MW; 9BB81992 CRC32;

Query Match 100.0%; Score 86; DB 4; Length 393;  
 Best Local Similarity 100.0%; Pred. No. 5.79e-08;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 127 SPALNKMFCQL 137

QY 1 SPALNKMFCQL 11

RESULT 2  
 ID Q15088 PRELIMINARY; PRT; 393 AA.  
 AC Q15088;

DT 01-NOV-1996 (TREMBlrel. 01, Created)  
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
 DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)  
 DE P53 TRANSFORMATION SUPPRESSOR (FRAGMENT).  
 GN P53.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Primates; Catarrhini; Hominidae; Homo.

RN SEQUENCE FROM N.A.  
RX MEDLINE: 92007731.  
RA FARRELL P.J., ALLAN G., SHANAHAN F., VOUSDEN K.H., CROOK T.;  
RT "p53 is frequently mutated in Burkitt's lymphoma cell lines.";  
RL EMO J. 10:2879-2887(1991).  
DR EMBL: X60016; CAA42631.1; -.  
DR HSSP: P04637; 1SAH.  
DR PFAM: PF00870; P53; 1.  
ET VARIANT 238 238 Y -> C.  
FT NON\_TER 393 393  
SQ SEQUENCE 393 AA; 43713 MW; A01E1523 CRC32;

Query Match 100.0%; Score 86; DB 4; Length 393;  
Best Local Similarity 100.0%; Pred. No. 5.79e-08;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 127 SPALNKMFCOL 137  
|||||  
QY 1 SPALNKMFCOL 11

RESULT 3  
ID Q15086 PRELIMINARY; PRT; 393 AA.  
AC Q15086;  
DT 01-NOV-1996 (TREMBlrel. 01, Created)  
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)  
DE P53 TRANSFORMATION SUPPRESSOR (FRAGMENT).  
GN P53.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 92007731.  
RA FARRELL P.J., ALLAN G., SHANAHAN F., VOUSDEN K.H., CROOK T.;  
RT "p53 is frequently mutated in Burkitt's lymphoma cell lines.";  
RL EMO J. 10:2879-2887(1991).  
DR EMBL: X60013; CAA42628.1; -.  
DR HSSP: P04637; 1SAH.  
DR PFAM: PF00870; P53; 1.  
ET VARIANT 246 246 T -> M.  
FT NON\_TER 393 393  
SQ SEQUENCE 393 AA; 43682 MW; 943B62A3 CRC32;

Query Match 100.0%; Score 86; DB 4; Length 393;  
Best Local Similarity 100.0%; Pred. No. 5.79e-08;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 127 SPALNKMFCOL 137  
|||||  
QY 1 SPALNKMFCOL 11

RESULT 4  
ID Q16810 PRELIMINARY; PRT; 393 AA.  
AC Q16810;  
DT 01-NOV-1996 (TREMBlrel. 01, Created)  
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)  
DE CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).  
GN P53.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 92007731.  
RA FARRELL P.J., ALLAN G., SHANAHAN F., VOUSDEN K.H., CROOK T.;  
RT "p53 is frequently mutated in Burkitt's lymphoma cell lines.";  
RL EMO J. 10:2879-2887(1991).  
CC -1- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT

CC PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL  
CC CYCLE REGULATION, IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY  
CC REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED  
CC FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF  
CC CYCLIN-DEPENDENT KINASES (BY SIMILARITY).  
CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
DR EMBL: X60020; CAA42635.1; -.  
DR HSSP: P04637; 1SAH.  
DR PFAM: PF00870; P53; 1.  
DR PFAM: PF00870; P53; 1.  
KW Anti-oncogene; DNA-binding; Transcription regulation; Activator;  
KW Nuclear protein; Phosphorylation.  
FT VARIANT 254 254 D -> N.  
FT NON\_TER 393 393  
SQ SEQUENCE 393 AA; 43714 MW; 5F914579 CRC32;

Query Match 100.0%; Score 86; DB 4; Length 393;  
Best Local Similarity 100.0%; Pred. No. 5.79e-08;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 127 SPALNKMFCOL 137  
|||||  
QY 1 SPALNKMFCOL 11

RESULT 5  
ID Q16807 PRELIMINARY; PRT; 393 AA.  
AC Q16807;  
DT 01-NOV-1996 (TREMBlrel. 01, Created)  
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)  
DE CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).  
GN P53.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 92007731.  
RA FARRELL P.J., ALLAN G., SHANAHAN F., VOUSDEN K.H., CROOK T.;  
RT "p53 is frequently mutated in Burkitt's lymphoma cell lines.";  
RL EMO J. 10:2879-2887(1991).  
CC -1- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT  
CC PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL  
CC CYCLE REGULATION, IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY  
CC REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED  
CC FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF  
CC CYCLIN-DEPENDENT KINASES (BY SIMILARITY).  
CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
DR EMBL: X60011; CAA42626.1; -.  
DR HSSP: P04637; 1SAH.  
DR PFAM: PF00870; P53; 1.  
DR PFAM: PF00870; P53; 1.  
KW Anti-oncogene; DNA-binding; Transcription regulation; Activator;  
KW Nuclear protein; Phosphorylation.  
FT VARIANT 193 193 R -> H.  
FT NON\_TER 393 393  
SQ SEQUENCE 393 AA; 43731 MW; 279BC9CB CRC32;

Query Match 100.0%; Score 86; DB 4; Length 393;  
Best Local Similarity 100.0%; Pred. No. 5.79e-08;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 127 SPALNKMFCOL 137  
|||||  
QY 1 SPALNKMFCOL 11

RESULT 6  
ID Q16808 PRELIMINARY; PRT; 393 AA.  
AC Q16808;  
DT 01-NOV-1996 (TREMBlrel. 01, Created)

```
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DE 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).
GN P53.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 92007731.
RA FARRELL P.J., ALLAN G., SHANAHAN F., VOUSDEN K.H., CROOK T.;
RT "P53 is frequently mutated in Burkitt's lymphoma cell lines.";
RL EMBL J. 10:2879-2887(1991).
CC -1- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT
CC PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL
CC CYCLE REGULATION, IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY
CC REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED
CC FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF
CC CYCLIN-DEPENDENT KINASES (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC EMBL; X60018; CAA42633.1; -.
DR HSSP; P04637; ISAH.
DR PROSITE; PS00348; P53; 1.
DR PFAM; PF00870; P53; 1.
KW Anti-oncogene; DNA-binding; Transcription regulation; Activator;
KW Nuclear protein; Phosphorylation.
FT VARIANT 163 163 H -> Y.
FT NON_TER 393
SQ SEQUENCE 393 AA; 43627 MW; AFD8A9E3 CRC32;

Query Match 100.0%; Score 86; DB 4; Length 393;
Best Local Similarity 100.0%; Pred. No. 5.79e-08;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 127 SPALNKMFCQL 137
QY 1 SPALNKMFCQL 11

RESULT 7
ID Q16535 PRELIMINARY; PRT; 393 AA.
AC Q16535;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1999 (TrEMBLrel. 01, Last sequence update)
DE 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE P53 TRANSFORMATION SUPPRESSOR (FRAGMENT).
GN P53.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 92007731.
RA FARRELL P.J., ALLAN G., SHANAHAN F., VOUSDEN K.H., CROOK T.;
RT "P53 is frequently mutated in Burkitt's lymphoma cell lines.";
RL EMBL; X60017; CAA42632.1; -.
DR EMBL; X60015; CAA42630.1; -.
DR HSSP; P04637; ISAH.
DR PFAM; PF00870; P53; 1.
FT VARIANT 248 248 Q -> R.
FT NON_TER 393
SQ SEQUENCE 393 AA; 43684 MW; 239818A9 CRC32;

Query Match 100.0%; Score 86; DB 4; Length 393;
Best Local Similarity 100.0%; Pred. No. 5.79e-08;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 127 SPALNKMFCQL 137
QY 1 SPALNKMFCQL 11

RESULT 8
ID Q16809 PRELIMINARY; PRT; 393 AA.
AC Q16809;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1999 (TrEMBLrel. 01, Last sequence update)
DE 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).
GN P53.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 92007731.
RA FARRELL P.J., ALLAN G., SHANAHAN F., VOUSDEN K.H., CROOK T.;
RT "P53 is frequently mutated in Burkitt's lymphoma cell lines.";
RL EMBL J. 10:2879-2887(1991).
CC -1- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT
CC PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL
CC CYCLE REGULATION, IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY
CC REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED
CC FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF
CC CYCLIN-DEPENDENT KINASES (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC EMBL; X60019; CAA42634.1; -.
DR HSSP; P04637; ISAH.
DR PROSITE; PS00348; P53; 1.
DR PFAM; PF00870; P53; 1.
KW Anti-oncogene; DNA-binding; Transcription regulation; Activator;
KW Nuclear protein; Phosphorylation.
FT VARIANT 213 213 Q -> R.
FT NON_TER 393
SQ SEQUENCE 393 AA; 43684 MW; CB70BD7F CRC32;

Query Match 100.0%; Score 86; DB 4; Length 393;
Best Local Similarity 100.0%; Pred. No. 5.79e-08;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 127 SPALNKMFCQL 137
QY 1 SPALNKMFCQL 11

RESULT 9
ID Q16848 PRELIMINARY; PRT; 393 AA.
AC Q16848;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1999 (TrEMBLrel. 01, Last sequence update)
DE 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE CELLULAR TUMOR ANTIGEN P53.
GN TP53.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 87089826.
RA HARRIS N., BRILL E., SHOHAT O., PROKOCIMER M., WOLF D., ARAI N.,
RA ROTTER V.;
RT "Molecular basis for heterogeneity of the human p53 protein.";
RL MOL. CELL. BIOL. 6:4650-4656(1986).
CC -1- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT
CC PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL
CC CYCLE REGULATION, IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY
CC REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED
CC FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF
CC CYCLIN-DEPENDENT KINASES (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC EMBL; M14694; AAA61211.1; -.
DR HSSP; P04637; ITSR.
DR PROSITE; PS00348; P53; 1.
DR PFAM; PF00870; P53; 1.
DR PRINTS; PR00386; P53SUPPRESSR.
```

KW Nuclear protein; Phosphorylation; Anti-oncogene; DNA-binding;  
KW Transcription regulation; Activator.  
SQ SEQUENCE 393 AA; 43723 MW; DA7D302F CRC32;

Query Match 100.0%; Score 86; DB 4; Length 393;  
Best Local Similarity 100.0%; Pred. No. 5.79e-08;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 127 SPALNKMFCOL 137  
|||||  
QY 1 SPALNKMFCOL 11

RESULT 10  
ID Q16811 PRELIMINARY; PRT; 393 AA.

AC Q16811;  
DT 01-NOV-1996 (TRENBLrel. 01, Created)  
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)  
DT 01-NOV-1999 (TRENBLrel. 12, Last annotation update)  
DE CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.  
RN [1]  
RN SEQUENCE FROM N.A.  
RX MEDLINE: 85126934.  
RA MATIASHEWSKI G., LAMB P., PIM D., PEACOCK J., CRAWFORD L.,  
RA BENCHIMOL S.;  
RT Isolation and characterization of a human p53 cDNA clone: expression  
of the human p53 gene.;  
RL EMBO J. 3:3257-3262(1984).  
RN [2]  
RN SEQUENCE FROM N.A.  
RX MEDLINE: 87064416.  
RA LAMB P., CRAWFORD L.;  
RT "Characterization of the human p53 gene.";  
RL Mol. Cell. Biol. 6:1379-1385(1986).  
CC -!- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT  
PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL  
CYCLE REGULATION, IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY  
REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED  
FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF  
CYCLIN-DEPENDENT KINASES (BY SIMILARITY).  
CC -!- SUBCELLULAR LOCATION: NUCLEAR.  
DR EMBL: M13121; AA59987.1; JOINED.  
DR EMBL: M13112; AA59987.1; JOINED.  
DR EMBL: M13113; AA59987.1; JOINED.  
DR EMBL: M13114; AA59987.1; JOINED.  
DR EMBL: M13115; AA59987.1; JOINED.  
DR EMBL: M13116; AA59987.1; JOINED.  
DR EMBL: M13117; AA59987.1; JOINED.  
DR EMBL: M13118; AA59987.1; JOINED.  
DR EMBL: M13119; AA59987.1; JOINED.  
DR EMBL: M13120; AA59987.1; JOINED.  
DR HSSP: P04637; ITSR.  
DR PROSITE: PS00348; P53; 1.  
DR PFAM: PF00870; P53; 1.  
KW Repeat: Tumor antigen; Anti-oncogene; DNA-binding;  
KW Transcription regulation; Activator; Nuclear protein; Phosphorylation.  
FT NON\_TER 393  
SQ SEQUENCE 393 AA; 43698 MW; 3EA71431 CRC32;

Query Match 100.0%; Score 86; DB 4; Length 393;  
Best Local Similarity 100.0%; Pred. No. 5.79e-08;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 127 SPALNKMFCOL 137  
|||||  
QY 1 SPALNKMFCOL 11

RESULT 11  
ID Q35873 PRELIMINARY; PRT; 205 AA.

AC Q35873;

DT 01-JAN-1998 (TRENBLrel. 05, Created)  
DT 01-JAN-1998 (TRENBLrel. 05, Last sequence update)  
DT 01-NOV-1999 (TRENBLrel. 12, Last annotation update)  
DE CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).  
GN P53.  
OS Cricetus griseus (Chinese hamster).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae; Cricetulus.  
RN [1]  
RN SEQUENCE FROM N.A.  
RA RAINALDI G., MARCHETTI S., CAPECCHI B., MENEVERI R., PIRAS A.,  
RA LEUZZI R.;  
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RN SEQUENCE FROM N.A.  
RA VATERONI L., MUSIO A., MENEVERI R., RAINALDI G.;  
RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT  
PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL  
CYCLE REGULATION, IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY  
REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED  
FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF  
CYCLIN-DEPENDENT KINASES (BY SIMILARITY).  
CC -!- SUBCELLULAR LOCATION: NUCLEAR.  
DR EMBL: U74487; AAB82420.1; -.  
DR HSSP: P04637; 1SAH.  
DR PROSITE: PS00348; P53; 1.  
DR PFAM: PF00870; P53; 1.  
KW Anti-oncogene; DNA-binding; Transcription regulation; Activator;  
KW Nuclear protein; Phosphorylation.  
FT NON\_TER 1  
FT NON\_TER 205  
SQ SEQUENCE 205 AA; 23122 MW; 680DDDDC CRC32;

Query Match 89.5%; Score 77; DB 11; Length 205;  
Best Local Similarity 81.8%; Pred. No. 1.25e-05;  
Matches 9; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 2 SPALNKMFCOL 12  
|||||  
QY 1 SPALNKMFCOL 11

RESULT 12  
ID P89004 PRELIMINARY; PRT; 238 AA.

AC P89004;  
DT 01-MAY-1997 (TRENBLrel. 03, Created)  
DT 01-MAY-1997 (TRENBLrel. 03, Last sequence update)  
DT 01-NOV-1999 (TRENBLrel. 12, Last annotation update)  
DE P53 (FRAGMENT).  
OS Mastomys natalensis papillomavirus (Mnpv).  
OC Viruses; dsDNA viruses, no RNA stage; Papovaviridae; Papillomavirus.  
RN [1]  
RN SEQUENCE FROM N.A.  
RC TISSUE-ECTOMA INDUCED BY LOXTIDINE.;  
RA LUQUE E.A., TANG L.H., MODLIN I.M.;  
RL Gastroenterology 0:0-0(0).  
DR EMBL: U48618; AAB41833.1; -.  
DR HSSP: P04637; 1YCS.  
DR PFAM: PF00870; P53; 1.  
FT NON\_TER 1  
SQ SEQUENCE 238 AA; 26704 MW; 097E01F9 CRC32;

Query Match 89.5%; Score 77; DB 14; Length 238;  
Best Local Similarity 81.8%; Pred. No. 1.25e-05;  
Matches 9; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 20 SPALNKMFCOL 30  
|||||  
QY 1 SPALNKMFCOL 11

RESULT 13  
ID P90332 PRELIMINARY; PRT; 286 AA.  
AC P90332;  
DT 01-MAY-1997 (TrEMBLrel. 03, Created)  
DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)  
DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)  
DE P53 (FRAGMENT).  
OS Mastomys natalensis papillomavirus (MnPV).  
OC Viruses; dsDNA viruses, no RNA stage; Papovaviridae; Papillomavirus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-SPONTANEOUS ECLOMAS;  
RA LUQUE E.A., TANG L.H., MODLIN I.M.;  
RL Gastroenterology 0:0-0(0).  
DR EMBL; U48619; AAB41834.1; -.  
DR HSSP; P04637; 1PET.  
DR PFAM; PF00870; P53; 1.  
FT NON\_TER 1  
SQ SEQUENCE 286 AA; 32247 MW; 5B5D3CAD CRC32;  
Query Match 89.5%; Score 77; DB 14; Length 286;  
Best Local Similarity 81.8%; Pred. No. 1.25e-05;  
Matches 9; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
Db 20 SP5LNKLFQOL 30  
||:|||||  
QY 1 SPALNRMFCOL 11

RESULT 14  
ID P89003 PRELIMINARY; PRT; 286 AA.  
AC P89003;  
DT 01-MAY-1997 (TrEMBLrel. 03, Created)  
DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)  
DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)  
DE P53 (FRAGMENT).  
OS Mastomys natalensis papillomavirus (MnPV).  
OC Viruses; dsDNA viruses, no RNA stage; Papovaviridae; Papillomavirus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA LUQUE E.A., TANG L.H., MODLIN I.M.;  
RL Gastroenterology 0:0-0(0).  
DR EMBL; U48617; AAB41832.1; -.  
DR HSSP; P04637; 1PET.  
DR PFAM; PF00870; P53; 1.  
FT NON\_TER 1  
SQ SEQUENCE 286 AA; 32287 MW; 30F7C9FA CRC32;  
Query Match 89.5%; Score 77; DB 14; Length 286;  
Best Local Similarity 81.8%; Pred. No. 1.25e-05;  
Matches 9; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
Db 20 SP5LNKLFQOL 30  
||:|||||  
QY 1 SPALNRMFCOL 11

RESULT 15  
ID P89002 PRELIMINARY; PRT; 378 AA.  
AC P89002;  
DT 01-MAY-1997 (TrEMBLrel. 03, Created)  
DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)  
DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)  
DE P53 (FRAGMENT).  
OS Mastomys natalensis papillomavirus (MnPV).  
OC Viruses; dsDNA viruses, no RNA stage; Papovaviridae; Papillomavirus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA LUQUE E.A., TANG L.H., MODLIN I.M.;  
RL Gastroenterology 0:0-0(0).  
DR EMBL; U48616; AAB41831.1; -.  
DR HSSP; P04637; 1PET.  
DR PFAM; PF00870; P53; 1.

FT NON\_TER 1  
SQ SEQUENCE 378 AA; 42062 MW; B4436760 CRC32;  
Query Match 89.5%; Score 77; DB 14; Length 378;  
Best Local Similarity 81.8%; Pred. No. 1.25e-05;  
Matches 9; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
Db 112 SP5LNKLFQOL 122  
||:|||||  
QY 1 SPALNRMFCOL 11  
Search completed: Sat Apr 15 00:31:48 2000  
Job time : 92 secs.

**THIS PAGE BLANK (USPTO)**

\*\*\*\*\*  
 WISE  
 \*\*\*\*\*  
 (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.  
 Copyright (c) 1993-1998 University of Edinburgh, U.K.  
 Distribution rights by Oxford Molecular Ltd

MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm  
 Run on: Sat Apr 15 00:35:26 2000; Maspar time 3.24 Seconds  
 Tabular output not generated. 65.788 Million cell updates/sec

Title: >US-08-452-843-16  
 Description: (1-9) from US08452843.pep  
 Perfect Score: 63  
 Sequence: 1 GTRVRAMAI 9

Scoring table: PAM 150  
 Gap 15

Searched: 188963 seqs, 23686106 residues

Post-processing: Minimum Match 0%  
 Listing first 45 summaries

Database: a:geneseq36  
 1:geneseq9

Statistics: Mean 15.963; Variance 44.724; scale 0.357

Pred. No. is the number of results predicted by chance to have a  
 score greater than or equal to the score of the result being printed,  
 and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	63	100.0	241	1 R51872	Human p53 amino acids	1.25e+00
2	63	100.0	253	1 W28484	Human p53 protein vari	1.25e+00
3	63	100.0	253	1 W28483	Human p53 protein vari	1.25e+00
4	63	100.0	319	1 W28495	Human p53 protein vari	1.25e+00
5	63	100.0	335	1 W28496	Human p53 protein vari	1.25e+00
6	63	100.0	335	1 W28498	Human p53 protein vari	1.25e+00
7	63	100.0	335	1 W28497	Human p53 protein vari	1.25e+00
8	63	100.0	353	1 W28493	Human p53 protein vari	1.25e+00
9	63	100.0	353	1 W28494	Human p53 protein vari	1.25e+00
10	63	100.0	354	1 R51874	Human p53 amino acids	1.25e+00
11	63	100.0	359	1 W13960	Chimeric p53 protein.	1.25e+00
12	63	100.0	361	1 W13961	Chimeric p53 protein.	1.25e+00
13	63	100.0	363	1 W28479	Human p53 protein vari	1.25e+00
14	63	100.0	363	1 W13972	Modified p53 variant p	1.25e+00
15	63	100.0	363	1 W28480	Human p53 protein vari	1.25e+00
16	63	100.0	374	1 W28482	Human p53 protein vari	1.25e+00
17	63	100.0	374	1 W28481	Human p53 protein vari	1.25e+00
18	63	100.0	381	1 W28490	Human p53 protein vari	1.25e+00
19	63	100.0	381	1 W28489	Human p53 protein vari	1.25e+00
20	63	100.0	393	1 Y03191	Amino acid sequence of	1.25e+00
21	63	100.0	393	1 W4270	Human p53 protein.	1.25e+00
22	63	100.0	393	1 W59218	Human p53 mutant 1.	1.25e+00
23	63	100.0	393	1 W69217	Human wild-type p53 pr	1.25e+00

## ALIGNMENTS

RESULT 1  
 ID R51872 standard; Protein; 241 AA.  
 AC R51872;  
 DT 18-NOV-1994 (first entry)  
 DE Human p53 amino acids 1-241.  
 KW Human nuclear phosphoprotein p53; tumour suppressor gene product;  
 KW anti-oncogene; cancer; tumour; antibody binding region; epitope.  
 OS Homo sapiens.  
 PN W09408241-A.  
 PD 14-APR-1994.  
 PF 30-SEP-1993; E026666.  
 PR 30-SEP-1992; DE-232823.  
 PA (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM STIFTUNG.  
 PI Klein R, Schranz P, Tessmer C, Volkman M, Zentgraf H;  
 DR WPI; 94-135732/16.  
 DR N-PSDB; Q62357.  
 PT Non-radioactive detection of p53 specific antibodies - by capture  
 PT on immobilised p53 or its fragments, then reaction with labelled  
 PT second antibody, for diagnosis of tumours and suitable for  
 PT screening  
 PS Claim 10; Page 17; 35pp; German.  
 CC Antibodies specific for p53 are detected by binding to immobilised  
 CC fragments of the p53 gene product containing the antibody-binding  
 CC region. Preferred fragments contain amino acids 1-241, 40-349,  
 CC 40-393, 66-241, 66-393, 237-349, 237-393 and esp. 9-33, 37-52 or  
 CC 368-386. See R51872-R51881 for sequences of these fragments.  
 SQ Sequence 241 AA;  
 Query Match 100.0%; Score 63; DB 1; Length 241;  
 Best Local Similarity 100.0%; Pred. No. 1.25e+00;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 154 GTRVRAMAI 162  
 |||||  
 QY 1 GTRVRAMAI 9  
 RESULT 2  
 ID W28484 standard; Protein; 253 AA.  
 AC W28484;  
 DT 25-NOV-1997 (first entry)  
 DE Human p53 protein variant V-367H.  
 KW Leucine zipper domain; LZD; oligomerisation domain; mutant; muten;  
 KW substitution; replacement; transactivation; viral protein VP16; HSV;  
 KW anti-oncogene; hyperproliferation; cancer; restenosis;  
 KW tumour suppression; apoptosis.  
 OS Chimeric - Homo sapiens.

Human p53 protein SEQ 1.25e+00  
 T284R modified human p 1.25e+00  
 Human wild-type p53 tu 1.25e+00  
 Human tumour-derived p 1.25e+00  
 Human p53 mutant R248Q 1.25e+00  
 Modified p53 variant p 1.25e+00  
 Modified p53 variant p 1.25e+00  
 Human p53 variant foun 1.25e+00  
 Human p53 mutant N239S 1.25e+00  
 Wild type p53 protein. 1.25e+00  
 Modified p53 variant p 1.25e+00  
 Human p53 tumour suppr 1.25e+00  
 Human tumour-derived p 1.25e+00  
 Human tumour-derived p 1.25e+00  
 Human p53 mutant R273C 1.25e+00  
 Human p53 protein vari 1.25e+00  
 Chimeric p53 protein. 1.25e+00  
 Chimeric p53 protein. 1.25e+00  
 Tumour suppressor prot 1.25e+00  
 Human p53 protein vari 1.25e+00

OS Chimeric - Herpes simplex virus.  
 OS Synthetic.  
 FH Key Location/Qualifiers  
 FT misc\_difference 189  
 FT /note= "Arg residue at position 182 of wild-type  
 FT p53 has been mutated to His"  
 PN WO9704092-A1.  
 PD 06-FEB-1997.  
 PF 17-JUL-1996; F01111.  
 PR 19-JUL-1995; FR-008729.  
 PA (RHON ) RHONE POULENC RORER SA.  
 PI Bracco L, Conseiller E;  
 DR WPI: 97-132633/12.  
 PT New p53 variants e.g. with oligomerisation domain replaced by  
 PT leucine zipper - useful for treating hyper-proliferative disorders,  
 PT esp. cancer and restenosis  
 PS Claim 32; Page -; 133pp; French.  
 CC Claimed variants of protein p53 have at least part of the p53  
 CC transactivation domain (amino acids 1-74) deleted and replaced by  
 CC the transactivating domain (TD) from herpes simplex virus viral  
 CC protein VP16 (amino acids 411-490). The present sequence is that of  
 CC a specifically claimed p53 variant designated V-367 and comprising  
 CC the VP16 TD and amino acids 75-367 of human wild-type p53 (but with  
 CC Arg182 replaced by His). The p53 variants are more active and more  
 CC stable tumour suppressors and apoptosis-inducing agents than wild-type  
 CC p53 and are active where the wild-type protein is not.  
 CC (Note: this sequence does not appear in the specification and has  
 CC been produced by modifying the given sequence of variant V-367).  
 CC Sequence 253 AA;  
 SQ

Query Match 100.0%; Score 63; DB 1; Length 253;  
 Best Local Similarity 100.0%; Pred. NO. 1.25e+00;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 162 GTRVRAMAI 170  
 |||||  
 QY 1 GTRVRAMAI 9

RESULT 3  
 ID W28483 standard; Protein; 253 AA.  
 AC W28483;  
 DT 25-NOV-1997 (first entry)  
 DE Human p53 protein variant V-367 encoded by pEC141.  
 KW Leucine zipper domain; LZD; oligomerisation domain; mutant; mutein;  
 KW substitution; replacement; transactivation; viral protein VP16; HSV;  
 KW anti-oncogene; hyperproliferation; cancer; restenosis;  
 KW tumour suppression; apoptosis.  
 OS Chimeric - Homo sapiens.  
 OS Synthetic.  
 PN WO9704092-A1.  
 PD 06-FEB-1997.  
 PF 17-JUL-1996; F01111.  
 PR 19-JUL-1995; FR-008729.  
 PA (RHON ) RHONE POULENC RORER SA.  
 PI Bracco L, Conseiller E;  
 DR WPI: 97-132633/12.  
 DR N-PSDB; T86217.  
 PT New p53 variants e.g. with oligomerisation domain replaced by  
 PT leucine zipper - useful for treating hyper-proliferative disorders,  
 PT esp. cancer and restenosis  
 PS Claim 32; Pages 80-81; 133pp; French.  
 CC Claimed variants of protein p53 have at least part of the p53  
 CC transactivation domain (amino acids 1-74) deleted and replaced by  
 CC the transactivating domain (TD) from herpes simplex virus viral  
 CC protein VP16 (amino acids 411-490). The present sequence is that of  
 CC a specifically claimed p53 variant designated V-367 and comprising  
 CC the VP16 TD with amino acids 75-367 of human wild-type p53. The p53  
 CC variants are more active and more stable tumour suppressors and  
 CC apoptosis-inducing agents than wild-type p53 and are active where  
 CC the wild-type protein is not.  
 CC Sequence 253 AA;  
 SQ

Query Match 100.0%; Score 63; DB 1; Length 253;  
 Best Local Similarity 100.0%; Pred. NO. 1.25e+00;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 162 GTRVRAMAI 170  
 |||||  
 QY 1 GTRVRAMAI 9

RESULT 4  
 ID W28495 standard; Protein; 319 AA.  
 AC W28495;  
 DT 25-NOV-1997 (first entry)  
 DE Human p53 protein variant 360-325 encoded by pEC178.  
 KW Leucine zipper domain; LZD; oligomerisation domain; mutant; mutein;  
 KW substitution; replacement; transactivation; viral protein VP16; HSV;  
 KW anti-oncogene; hyperproliferation; cancer; restenosis;  
 KW tumour suppression; apoptosis.  
 OS Homo sapiens.  
 OS Synthetic.  
 PN WO9704092-A1.  
 PD 06-FEB-1997.  
 PF 17-JUL-1996; F01111.  
 PR 19-JUL-1995; FR-008729.  
 PA (RHON ) RHONE POULENC RORER SA.  
 PI Bracco L, Conseiller E;  
 DR WPI: 97-132633/12.  
 DR N-PSDB; T86223.  
 PT New p53 variants e.g. with oligomerisation domain replaced by  
 PT leucine zipper - useful for treating hyper-proliferative disorders,  
 PT esp. cancer and restenosis  
 PS Claim 38; Pages 92-94; 133pp; French.  
 CC Claimed variants of protein p53 have at least part of the  
 CC oligomerisation domain deleted and replaced by a leucine zipper  
 CC domain. The mutants preferably also have at least part of the p53  
 CC transactivation domain (amino acids 1-74) deleted and replaced by  
 CC the domain 325-360 of p53. The present sequence is that of a  
 CC specifically claimed p53 variant designated 360-325 and comprising  
 CC the 325-360 domain, amino acids 75-325 of human wild-type p53 and a  
 CC leucine zipper domain at the C-terminal. The p53 variants are  
 CC more active and more stable tumour suppressors and apoptosis-inducing  
 CC agents than wild-type p53 and are active where the wild-type protein  
 CC is not, i.e. they are not inactivated by dominant negative or oncogenic  
 CC mutants, nor by other cellular proteins (because the leucine zipper  
 CC domain prevents formation of inactive mixed oligomers).  
 CC Sequence 319 AA;  
 SQ

Query Match 100.0%; Score 63; DB 1; Length 319;  
 Best Local Similarity 100.0%; Pred. NO. 1.25e+00;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 118 GTRVRAMAI 126  
 |||||  
 QY 1 GTRVRAMAI 9

RESULT 5  
 ID W28496 standard; Protein; 319 AA.  
 AC W28496;  
 DT 25-NOV-1997 (first entry)  
 DE Human p53 protein variant 360-325H.  
 KW Leucine zipper domain; LZD; oligomerisation domain; mutant; mutein;  
 KW substitution; replacement; transactivation; viral protein VP16; HSV;  
 KW anti-oncogene; hyperproliferation; cancer; restenosis;  
 KW tumour suppression; apoptosis.  
 OS Homo sapiens.  
 OS Synthetic.  
 PN WO9704092-A1.  
 PD 06-FEB-1997.  
 PF 17-JUL-1996; F01111.  
 PR 19-JUL-1995; FR-008729.  
 PA (RHON ) RHONE POULENC RORER SA.  
 PI Bracco L, Conseiller E;  
 DR WPI: 97-132633/12.  
 DR N-PSDB; T86217.  
 PT New p53 variants e.g. with oligomerisation domain replaced by  
 PT leucine zipper - useful for treating hyper-proliferative disorders,  
 PT esp. cancer and restenosis  
 PS Claim 32; Pages 80-81; 133pp; French.  
 CC Claimed variants of protein p53 have at least part of the p53  
 CC transactivation domain (amino acids 1-74) deleted and replaced by  
 CC the transactivating domain (TD) from herpes simplex virus viral  
 CC protein VP16 (amino acids 411-490). The present sequence is that of  
 CC a specifically claimed p53 variant designated V-367 and comprising  
 CC the VP16 TD with amino acids 75-367 of human wild-type p53. The p53  
 CC variants are more active and more stable tumour suppressors and  
 CC apoptosis-inducing agents than wild-type p53 and are active where  
 CC the wild-type protein is not.  
 CC Sequence 319 AA;  
 SQ

Query Match 100.0%; Score 63; DB 1; Length 319;  
 Best Local Similarity 100.0%; Pred. NO. 1.25e+00;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 118 GTRVRAMAI 126  
 |||||  
 QY 1 GTRVRAMAI 9

RESULT 5  
 ID W28496 standard; Protein; 319 AA.  
 AC W28496;  
 DT 25-NOV-1997 (first entry)  
 DE Human p53 protein variant 360-325H.  
 KW Leucine zipper domain; LZD; oligomerisation domain; mutant; mutein;  
 KW substitution; replacement; transactivation; viral protein VP16; HSV;  
 KW anti-oncogene; hyperproliferation; cancer; restenosis;  
 KW tumour suppression; apoptosis.  
 OS Homo sapiens.  
 OS Synthetic.  
 PN WO9704092-A1.  
 PD 06-FEB-1997.  
 PF 17-JUL-1996; F01111.  
 PR 19-JUL-1995; FR-008729.  
 PA (RHON ) RHONE POULENC RORER SA.  
 PI Bracco L, Conseiller E;  
 DR WPI: 97-132633/12.  
 DR N-PSDB; T86217.  
 PT New p53 variants e.g. with oligomerisation domain replaced by  
 PT leucine zipper - useful for treating hyper-proliferative disorders,  
 PT esp. cancer and restenosis  
 PS Claim 32; Pages 80-81; 133pp; French.  
 CC Claimed variants of protein p53 have at least part of the p53  
 CC transactivation domain (amino acids 1-74) deleted and replaced by  
 CC the transactivating domain (TD) from herpes simplex virus viral  
 CC protein VP16 (amino acids 411-490). The present sequence is that of  
 CC a specifically claimed p53 variant designated V-367 and comprising  
 CC the VP16 TD with amino acids 75-367 of human wild-type p53. The p53  
 CC variants are more active and more stable tumour suppressors and  
 CC apoptosis-inducing agents than wild-type p53 and are active where  
 CC the wild-type protein is not.  
 CC Sequence 319 AA;  
 SQ



CC sequence (Gly4Ser)3, and with a leucine zipper domain at the C-terminal. The p53 variants are more active and more stable tumour suppressors and apoptosis-inducing agents than wild-type p53 and are active where the wild-type protein is not, i.e. they are not inactivated by dominant negative or oncogenic mutants, nor by other cellular proteins (because the leucine zipper domain prevents formation of inactive mixed oligomers).

CC (Note: this sequence does not appear in the specification and has been produced by modifying the given sequence of variant 360h-325).

CC Sequence 335 AA;

CC Query Match 100.08; Score 63; DB 1; Length 335;

CC Best Local Similarity 100.08; Pred. No. 1.25e+00;

CC Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0

CC Sequence 335 AA;

CC Db 134 GTRVRAMAI 142

CC | | | | | | | | | |

CC QY 1 GTRVRAMAI 9

CC | | | | | | | | | |

CC RESULT 7

CC ID W28497 standard; Protein; 335 AA.

CC AC W28497;

CC DT 23-NOV-1997 (first entry)

CC DE Human p53 protein variant 360h-325 encoded by pBC179.

CC LE Leucine zipper domain; LZD; oligomerisation domain; mutant; mutein;

CC KW substitution; replacement; transactivation; hinge region;

CC KW anti-oncogene; hyperproliferation; cancer; restenosis;

CC KW tumour suppression; apoptosis.

CC OS Homo sapiens.

CC OS Synthetic.

CC FH Key

CC FT region

CC FT 39..53

CC FT /label= hinge

CC PN W09704092-A1.

CC PD 08-FEB-1997.

CC PF 17-JUL-1996; F01111.

CC PR 13-JUL-1995; FR-008729.

CC PA (RHON ) RHONE POULENC RORER SA.

CC PI Bracco L, Conseiller E;

CC DR WPI: 97-132633/12.

CC DR N-PSDB: T86224.

CC DR New p53 variants e.g. with oligomerisation domain replaced by

CC PT leucine zipper - useful for treating hyper-proliferative disorders,

CC PT esp. cancer and restenosis

CC PS Claim 39; Pages 94-95; 133pp; French.

CC CC Claimed variants of protein p53 have at least part of the

CC CC oligomerisation domain deleted and replaced by a leucine zipper

CC CC domain. The mutants preferably also have at least part of the p53

CC CC transactivation domain (amino acids 1-74) deleted and replaced by

CC CC the domain 325-360 of p53. The present sequence is that of a

CC CC specifically claimed p53 variant designated 360h-325 and comprising

CC CC the 325-360 domain, separated from amino acids 75-325 of human

CC CC wild-type p53 by a synthetic hinge sequence (Gly4Ser)3, and with a

CC CC leucine zipper domain at the C-terminal. The p53 variants are

CC CC more active and more stable tumour suppressors and apoptosis-inducing

CC CC agents than wild-type p53 and are active where the wild-type protein

CC CC is not, i.e. they are not inactivated by dominant negative or oncogenic

CC CC mutants, nor by other cellular proteins (because the leucine zipper

CC CC domain prevents formation of inactive mixed oligomers).

CC Sequence 335 AA;

CC Query Match 100.08; Score 63; DB 1; Length 335;

CC Best Local Similarity 100.08; Pred. No. 1.25e+00;

CC Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0

CC Db 134 GTRVRAMAI 142

CC | | | | | | | | | |

CC QY 1 GTRVRAMAI 9

CC | | | | | | | | | |

CC RESULT 8

CC ID W28493 standard; Protein; 353 AA.

AC W28493;  
 DT 25-NOV-1997 (first entry)  
 DE Human p53 protein variant 393-325 encoded by pEC177.  
 KW Leucine zipper domain; LZD; oligomerisation domain; mutant; mutein;  
 KW substitution; replacement; transactivation; viral protein VP16; HSV;  
 KW anti-oncogene; hyperproliferation; cancer; restenosis;  
 KW tumour suppression; apoptosis.  
 OS Homo sapiens.  
 PN WO9704092-A1.  
 PD 06-FEB-1997.  
 PF 17-JUL-1996; F01111.  
 PR 19-JUL-1995; FR-008729.  
 PA (RHON ) RHONE POULENC RORER SA.  
 PI Bracco L, Conseiller E;  
 DR WPI: 97-132633/12.  
 DR N-PSDB: 786222.  
 FT New p53 variants e.g. with oligomerisation domain replaced by  
 FT leucine zipper - useful for treating hyper-proliferative disorders,  
 FT esp. cancer and restenosis  
 PS Claim 37: Pages 90-92; 133pp; French.  
 CC Claimed variants of protein p53 have at least part of the  
 CC oligomerisation domain deleted and replaced by a leucine zipper  
 CC domain. The mutants preferably also have at least part of the p53  
 CC transactivation domain (amino acids 1-74) deleted and replaced by  
 CC the domain 325-393 of p53. The present sequence is that of  
 CC a specifically claimed p53 variant designated 393-325 and comprising  
 CC the 325-393 domain, amino acids 75-325 of human wild-type p53 and a  
 CC leucine zipper domain at the C-terminal. The p53 variants are  
 CC more active and more stable tumour suppressors and apoptosis-inducing  
 CC agents than wild-type p53 and are active where the wild-type protein  
 CC is not, i.e. they are not inactivated by dominant negative or oncogenic  
 CC mutants, nor by other cellular proteins (because the leucine zipper  
 CC domain prevents formation of inactive mixed oligomers).  
 SQ Sequence 353 AA;  
 Query Match 100.0%; Score 63; DB 1; Length 353;  
 Best Local Similarity 100.0%; Pred. No. 1.25e+00;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 152 GTRVRAMAI 160  
 QY 1 GTRVRAMAI 9  
 |||||  
 RESULT 9  
 ID W28494 standard; Protein; 353 AA.  
 AC W28494;  
 DT 25-NOV-1997 (first entry)  
 DE Human p53 protein variant 393-325H.  
 KW Leucine zipper domain; LZD; oligomerisation domain; mutant; mutein;  
 KW substitution; replacement; transactivation; viral protein VP16; HSV;  
 KW anti-oncogene; hyperproliferation; cancer; restenosis;  
 KW tumour suppression; apoptosis.  
 OS Homo sapiens.  
 PN WO9704092-A1.  
 PD 06-FEB-1997.  
 PF 17-JUL-1996; F01111.  
 PR 19-JUL-1995; FR-008729.  
 PA (RHON ) RHONE POULENC RORER SA.  
 PI Bracco L, Conseiller E;  
 DR WPI: 97-132633/12.  
 FT New p53 variants e.g. with oligomerisation domain replaced by  
 FT leucine zipper - useful for treating hyper-proliferative disorders,  
 FT esp. cancer and restenosis  
 PS Claim 37: Page: 133pp; French.  
 CC Claimed variants of protein p53 have at least part of the  
 CC oligomerisation domain deleted and replaced by a leucine zipper

CC domain. The mutants preferably also have at least part of the p53  
 CC transactivation domain (amino acids 1-74) deleted and replaced by  
 CC the domain 325-393 of p53. The present sequence is that of a  
 CC specifically claimed p53 variant designated 393-325H and comprising  
 CC the 325-393 domain, amino acids 75-325 of human wild-type p53 (but with  
 CC Arg182 replaced by His) and a leucine zipper domain at the C-terminal.  
 CC The p53 variants are more active and more stable tumour suppressors  
 CC and apoptosis-inducing agents than wild-type p53 and are active where  
 CC the wild-type protein is not, i.e. they are not inactivated by dominant  
 CC negative or oncogenic mutants, nor by other cellular proteins (because  
 CC the leucine zipper domain prevents formation of inactive mixed  
 CC oligomers).  
 CC (Note: this sequence does not appear in the specification and has  
 CC been produced by modifying the given sequence of variant 393-325).  
 SQ Sequence 353 AA;  
 Query Match 100.0%; Score 63; DB 1; Length 353;  
 Best Local Similarity 100.0%; Pred. No. 1.25e+00;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 152 GTRVRAMAI 160  
 QY 1 GTRVRAMAI 9  
 |||||  
 RESULT 10  
 ID R51874 standard; Protein; 354 AA.  
 AC R51874;  
 DT 18-NOV-1994 (first entry)  
 DE Human p53 amino acids 40-393.  
 KW Human nuclear phosphoprotein p53; tumour suppressor gene product;  
 KW anti-oncogene; cancer; tumour; antibody binding region; epitope.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT misc\_difference 234  
 FT /note= "Arg corresponds to a CAT codon"  
 PN W09408241-A.  
 PD 14-APR-1994.  
 PF 30-SEP-1993; E02666.  
 PR 30-SEP-1992; DE-232823.  
 PA (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM STIFTUNG.  
 PI Klein R, Schranz P, Tessmer C, Volkman M, Zentgraf H;  
 DR WPI: 94-135732/16.  
 DR N-PSDB: Q62359.  
 FT Non-radioactive detection of p53 specific antibodies - by capture  
 FT on immobilised p53 or its fragments, then reaction with labelled  
 FT second antibody, for diagnosis of tumours and suitable for  
 FT screening  
 PT Claim 10; Page 18; 35pp; German.  
 CC Antibodies specific for p53 are detected by binding to immobilised  
 CC fragments of the p53 gene product containing the antibody-binding  
 CC region. Preferred fragments contain amino acids 1-241, 40-349,  
 CC 40-393, 66-241, 66-393, 237-349, 237-393 and esp. 9-33, 37-52 or  
 CC 368-386. See R51872-R51881 for sequences of these fragments.  
 SQ Sequence 354 AA;  
 Query Match 100.0%; Score 63; DB 1; Length 354;  
 Best Local Similarity 100.0%; Pred. No. 1.25e+00;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 115 GTRVRAMAI 123  
 QY 1 GTRVRAMAI 9  
 |||||  
 RESULT 11  
 ID W13960 standard; Protein; 359 AA.  
 AC W13960;  
 DT 25-JUN-1997 (first entry)  
 DE Chimeric p53 protein.  
 KW p53; tumour suppressor; cancer; therapy; cell proliferation;  
 KW apoptosis; protein engineering; GCNA; DNA binding.  
 OS Chimeric Homo sapiens;

OS Chimeric synthetic.  
FH Key Location/Qualifiers  
FT region 1..323  
FT /label= p53wt  
FT /note= "amino acids 1-323 of wild-type p53"  
FT 324..326  
FT /label= Linker  
FT region 327..359  
FT /label= GCN4  
FT /note= "amino acids 249-281 of GCN4 LZ variant"  
PN W09710843-A1.  
PD 27-MAR-1997.  
PF 20-SEP-1996; U15188.  
PR 22-SEP-1995; US-004802.  
PR 21-AUG-1996; US-697221.  
PA (WIST-) WISTAR INST ANATOMY & BIOLOGY.  
PI Halazonetis TD;  
DR WPI; 97-202618/18.  
PT R284K modified p53 protein having DNA binding ability - useful in treatment of cancer  
PS Disclosure; Refer to Page 8; 82pp; English.  
CC Chimeric p53 constructs (W13956-67) comprise N-terminal portions of human wild-type p53 tumour suppressor (see also W13948) linked to a C-terminal portion of the LZ variant (see also W13955) of GCN4 and, in some cases, the C-terminal portion of wild-type p53. The chimeric proteins have DNA binding activity and can replace lost or insufficient p53 function, providing the means for pharmacological rescue of p53 function in cancer patients. Nucleic acids coding for modified p53 constructs can be used for cancer gene therapy.  
SQ Sequence 359 AA;  
  
Query Match 100.0%; Score 63; DB 1; Length 359;  
Best Local Similarity 100.0%; Pred. No. 1.25e+00;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Db 154 GTRVRAMAI 162  
QY 1 GTRVRAMAI 9  
  
RESULT 12  
ID W13961 standard; Protein; 361 AA.  
AC W13961;  
DT 25-JUN-1997 (first entry)  
DE Chimeric p53 protein.  
KW p53; tumour suppressor; cancer; therapy; cell proliferation; apoptosis; protein engineering; GCN4; DNA binding.  
KW Chimeric Homo sapiens;  
OS Chimeric synthetic.  
FH Key Location/Qualifiers  
FT region 1..323  
FT /label= p53wt  
FT /note= "amino acids 1-323 of wild-type p53"  
FT 324..329  
FT /label= Linker  
FT region 330..361  
FT /label= GCN4  
FT /note= "amino acids 250-281 of GCN4 LZ variant"  
PN W09710843-A1.  
PD 27-MAR-1997.  
PF 20-SEP-1996; U15188.  
PR 22-SEP-1995; US-004802.  
PR 21-AUG-1996; US-697221.  
PA (WIST-) WISTAR INST ANATOMY & BIOLOGY.  
PI Halazonetis TD;  
DR WPI; 97-202618/18.  
PT R284K modified p53 protein having DNA binding ability - useful in treatment of cancer  
PS Disclosure; Refer to Page 8; 82pp; English.  
CC Chimeric p53 constructs (W13956-67) comprise N-terminal portions of human wild-type p53 tumour suppressor (see also W13948) linked to a C-terminal portion of the LZ variant (see also W13955) of

CC GCN4 and, in some cases, the C-terminal portion of wild-type p53. The chimeric proteins have DNA binding activity and can replace lost or insufficient p53 function, providing the means for pharmacological rescue of p53 function in cancer patients. Nucleic acids coding for modified p53 constructs can be used for cancer gene therapy.  
SQ Sequence 361 AA;  
  
Query Match 100.0%; Score 63; DB 1; Length 361;  
Best Local Similarity 100.0%; Pred. No. 1.25e+00;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Db 154 GTRVRAMAI 162  
QY 1 GTRVRAMAI 9  
  
RESULT 13  
ID W28479 standard; Protein; 363 AA.  
AC W28479;  
DT 25-NOV-1997 (first entry)  
DE Human p53 protein variant V-325 encoded by pEC114.  
KW Leucine zipper domain; LZD; oligomerisation domain; mutant; muten; substitution; replacement; transactivation; viral protein VP16; HSV; anti-oncogene; hyperproliferation; cancer; restenosis;  
KW tumour suppression; apoptosis.  
OS Chimeric - Homo sapiens.  
OS Chimeric - Herpes simplex virus.  
OS Synthetic.  
PN W09704092-A1.  
PD 06-FEB-1997.  
PF 17-JUL-1996; FO1111.  
PR 19-JUL-1995; FR-008729.  
PA (RHON ) RHONE POULENC RORER SA.  
PI Bracco L; Conseiller E;  
DR WPI; 97-132833/12.  
DR N-PSDB; T86215.  
DR New p53 variants e.g. with oligomerisation domain replaced by leucine zipper - useful for treating hyper-proliferative disorders, esp. cancer and restenosis  
PT Claim 30; Pages 76-78; 133pp; French.  
PS Claimed variants of protein p53 have at least part of the oligomerisation domain deleted and replaced by a leucine zipper domain. The mutants preferably also have at least part of the p53 transactivation domain (amino acids 1-74) deleted and replaced by the transactivating domain (TD) from herpes simplex virus viral protein VP16 (amino acids 411-490). The present sequence is that of a specifically claimed p53 variant designated V-325 and comprising the VP16 TD, amino acids 75-325 of human wild-type p53 and a leucine zipper domain at the C-terminal. The p53 variants are more active and more stable tumour suppressors and apoptosis-inducing agents than wild-type p53 and are active where the wild-type protein is not, i.e. they are not inactivated by dominant negative or oncogenic mutants, nor by other cellular proteins (because the leucine zipper domain prevents formation of inactive mixed oligomers).  
SQ Sequence 363 AA;  
  
Query Match 100.0%; Score 63; DB 1; Length 363;  
Best Local Similarity 100.0%; Pred. No. 1.25e+00;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Db 162 GTRVRAMAI 170  
QY 1 GTRVRAMAI 9  
  
RESULT 14  
ID W13972 standard; Protein; 363 AA.  
AC W13972;  
DT 25-JUN-1997 (first entry)  
DE Modified p53 variant p53Q248del364-393.  
KW p53; tumour suppressor; cancer; therapy; cell proliferation; apoptosis; protein engineering; DNA binding.

OS Synthetic.  
 PN WO9710843-A1.  
 PD 27-MAR-1997.  
 PF 20-SEP-1996; U15188.  
 PR 22-SEP-1995; US-004802.  
 PR 21-AUG-1996; US-697221.  
 PA (WIST-) WISTAR INST ANATOMY & BIOLOGY.  
 PI Halazonetis TD;  
 DR WPI: 97-202618/18.  
 PT R284K modified p53 protein having DNA binding ability - useful in  
 treatment of cancer.  
 PS Example 1; 53-54; 82pp; English.  
 CC Modified p53 variant p53Q248del364-393 (W13972) has the tumour-  
 derived glutamine 248 mutation (see also W13951) and a deletion  
 of the C-terminal 30 amino acids of wild-type p53 (see also  
 CC W13948). Gln248 is a Class I p53 tumour mutation that affects DNA  
 binding. The C-terminal deletion, introduced by site-directed  
 CC mutagenesis of p53 DNA, activates the DNA binding of the p53  
 CC tumour mutant. This provides the means for pharmacological rescue  
 CC of p53 function in cancer patients. Other modified p53 constructs  
 CC (W13949-50, W13953-54, W13968-77) have also been produced. Nucleic  
 CC acids coding for modified p53 can be used for cancer gene therapy.  
 SQ Sequence 363 AA;

Query Match 100.0%; Score 63; DB 1; Length 363;  
 Best Local Similarity 100.0%; Pred. No. 1.25e+00;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 162 GTRVRAMAI 170  
 QY 1 GTRVRAMAI 9  
 |||||

Search completed: Sat Apr 15 00:36:02 2000  
 Job time : 36 secs.

Query Match 100.0%; Score 63; DB 1; Length 363;  
 Best Local Similarity 100.0%; Pred. No. 1.25e+00;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 154 GTRVRAMAI 162  
 QY 1 GTRVRAMAI 9  
 |||||

RESULT 15  
 ID W28480 standard; Protein; 363 AA.  
 AC W28480;  
 DT 25-NOV-1997 (first entry)  
 DE Human p53 protein variant V-325H.  
 KW Leucine zipper domain; LZD; oligomerisation domain; mutant; mutein;  
 KW substitution; replacement; transactivation; viral protein VP16; HSV;  
 KW anti-oncogene; hyperproliferation; cancer; restenosis;  
 KW tumour suppression; apoptosis.  
 OS Chimeric - Homo sapiens.  
 OS Chimeric - Herpes simplex virus.  
 OS Synthetic.  
 FH Key Location/Qualifiers  
 FT misc\_difference 189  
 FT /note= "Arg residue at position 182 of wild-type  
 FT p53 has been mutated to His"  
 PN WO9704092-A1.  
 PD 06-FEB-1997.  
 PF 17-JUL-1996; F01111.  
 PR 19-JUL-1995; PR-008729.  
 PA (RHON ) RHONE POULENC RORER SA.  
 PI Bracco L, Conseiller E;  
 DR WPI: 97-132633/12.  
 PT New p53 variants e.g. with oligomerisation domain replaced by  
 PT leucine zipper - useful for treating hyper-proliferative disorders,  
 PT esp. cancer and restenosis  
 PS Claim 30; Page : 133pp; French.  
 CC Claimed variants of protein p53 have at least part of the  
 CC oligomerisation domain deleted and replaced by a leucine zipper  
 CC domain. The mutants preferably also have at least part of the p53  
 CC transactivation domain (amino acids 1-74) deleted and replaced by  
 CC the transactivating domain (TD) from herpes simplex virus viral  
 CC protein VP16 (amino acids 411-490). The present sequence is that of  
 CC a specifically claimed p53 variant designated V-325H and comprising  
 CC the VP16 TD, amino acids 75-325 of human wild-type p53 (but with  
 CC Arg182 replaced by His) and a leucine zipper domain at the C-terminal.  
 CC The p53 variants are more active and more stable tumour suppressors  
 CC and apoptosis-inducing agents than wild-type p53 and are active where  
 CC the wild-type protein is not, i.e. they are not inactivated by dominant  
 CC negative or oncogenic mutants, nor by other cellular proteins (because

CC the leucine zipper domain prevents formation of inactive mixed  
 CC oligomers).  
 CC (Note: this sequence does not appear in the specification and has  
 CC been produced by modifying the given sequence of variant V-325).  
 SQ Sequence 363 AA;

Query Match 100.0%; Score 63; DB 1; Length 363;  
 Best Local Similarity 100.0%; Pred. No. 1.25e+00;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 162 GTRVRAMAI 170  
 QY 1 GTRVRAMAI 9  
 |||||

\*\*\*\*\*  
M P E R L H  
(TM)  
\*\*\*\*\*

Release 3.1A John F. Collins, BioComputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MParch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Sat Apr 15 00:36:19 2000; MasPar time 3.23 Seconds  
Tabular output not generated.  
111.755 Million cell updates/sec

Title: >US-08-452-843-16  
Description: (1-9) from US08452843.pep  
Perfect Score: 63  
Sequence: 1 GTRVRAMAI 9

Scoring table: PAM 150  
Gap 15

Searched: 122810 seqs, 40068593 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: pir62  
1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 22.750; Variance 26.719; scale 0.851

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	63	100.0	386	2 S51648	cellular tumor antigen	2.91e-03
2	63	100.0	391	2 S02192	cellular tumor antigen	2.91e-03
3	63	100.0	391	2 JC6193	tumor suppressor p53	2.91e-03
4	63	100.0	393	2 JC6176	tumor suppressor prot	2.91e-03
5	63	100.0	393	1 DNH053	cellular tumor antigen	2.91e-03
6	63	100.0	396	2 JH0633	cellular tumor antigen	2.91e-03
7	60	95.2	381	2 S38824	cellular tumor antigen	1.56e-02
8	60	95.2	390	1 DNMS53	cellular tumor antigen	1.56e-02
9	60	95.2	393	2 S06594	cellular tumor antigen	1.56e-02
10	54	85.7	564	2 B69137	sensory transduction	3.96e-01
11	53	84.1	567	2 H69145	sensory transduction	6.66e-01
12	52	82.5	248	2 S38658	DNA-directed RNA poly	1.11e+00
13	50	79.4	386	2 JC4865	contractile tail shea	3.05e+00
14	50	79.4	386	2 JC5191	tail sheath protein -	3.05e+00
15	49	77.8	387	1 XUEC	acetyl-CoA C-acyltran	5.00e+00
16	49	77.8	495	2 A69160	sensory transduction	5.00e+00
17	48	76.2	209	2 G70689	hypothetical protein	8.13e+00
18	48	76.2	383	2 F70903	probable adhe protein	8.13e+00
19	47	74.6	244	2 D69758	hypothetical protein	1.31e+01
20	47	74.6	520	2 C70776	probable export prote	1.31e+01
21	47	74.6	573	2 D64321	DNA ligase (ATP) (EC	1.31e+01
22	46	73.0	192	2 JC5876	early light-inducible	2.11e+01
23	46	73.0	229	2 JY0043	SpoU protein - Escher	2.11e+01

24 46 73.0 408 1 F41858 biphenyl dioxygenase 2.11e+01  
25 46 73.0 408 1 E42409 biphenyl dioxygenase 2.11e+01  
26 46 73.0 410 1 D36516 ribulose biphosphate 3.36e+01  
27 45 71.4 189 1 RKXHS ribulose biphosphate 3.36e+01  
28 45 71.4 339 2 H71265 hypothetical protein 3.36e+01  
29 45 71.4 504 2 A57215 glial cells missing ( 3.36e+01  
30 44 69.8 92 2 S43106 orfl protein - Yersin 5.30e+01  
31 44 69.8 179 1 R5B55 ribosomal protein L5 5.30e+01  
32 44 69.8 187 2 G70643 probable rplE protein 5.30e+01  
33 44 69.8 191 2 S29884 ribosomal protein L5 5.30e+01  
34 44 69.8 277 2 E69301 hypothetical protein 5.30e+01  
35 44 69.8 369 2 S53722 farnesyltransferase 5.30e+01  
36 44 69.8 382 2 B69877 sulfate adenylitrans 5.30e+01  
37 44 69.8 443 2 D69306 conserved hypothetical 5.30e+01  
38 44 69.8 450 2 B69198 UDP-N-acetylmuramyl t 5.30e+01  
39 44 69.8 469 2 D64661 ATP synthase F1, subu 5.30e+01  
40 44 69.8 469 2 D71855 ATP synthase F1, chai 5.30e+01  
41 44 69.8 486 2 I39523 dehydroshikimate dehy 5.30e+01  
42 44 69.8 501 2 S22669 hypothetical protein 5.30e+01  
43 44 69.8 661 2 H64876 probable membrane pro 5.30e+01  
44 44 69.8 707 2 T02835 long chain fatty acyl 5.30e+01  
45 44 69.8 1953 2 S63244 BNL1 protein - yeast 5.30e+01

ALIGNMENTS

RESULT 1  
ENTRY S51648 #type complete  
TITLE cellular tumor antigen p53 - bovine  
ALTERNATE\_NAMES tumor-suppressor protein p53  
ORGANISM #formal\_name Bos primigenius taurus #common\_name cattle  
DATE 07-May-1995 #sequence\_revision 01-Sep-1995 #text\_change 08-Sep-1997

ACCESSIONS S51648  
REFERENCE S51648  
#authors Dequiedt, F.; Willens, L.; Burny, A.; Kettmann, R.  
#submission Submitted to the EMBL Data Library, September 1994  
#description Nucleotide sequence of the ovine p53 tumor-suppressor gene  
#accession S51648 CDNA and its genomic organisation.

##status preliminary  
##molecule\_type mRNA  
##residues 1-386 #label DEQ  
##cross-references EMBL:X81704; NID:g602332; PID:g602333  
CLASSIFICATION #superfamily cellular tumor antigen p53  
KEYWORDS apoptosis; cell division control; DNA binding; homotetramer; phosphoprotein; transcription regulation; tumor suppressor; zinc

FEATURE 168,171,231,235 #binding\_site zinc (Cys, His, Cys, Cys) #status predicted  
385 #binding\_site phosphoryl-RNA (Ser) (covalent) #status predicted

SUMMARY #length 386 #molecular-weight 43255 #checksum 7025  
Query Match 100.0%; Score 63; DB 2; Length 386;  
Best Local Similarity 100.0%; Pred. No. 2.91e-03;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 146 GTRVRAMAI 154  
|||||  
QY 1 GTRVRAMAI 9

RESULT 2  
ENTRY S02192 #type complete  
TITLE cellular tumor antigen p53 - rat  
ALTERNATE\_NAMES gene p53 protein; nuclear oncoprotein p53  
ORGANISM #formal\_name Rattus norvegicus #common\_name Norway rat  
DATE 18-Oct-1989 #sequence\_revision 18-Oct-1989 #text\_change 17-Mar-1999  
ACCESSIONS S02192; S41149  
REFERENCE S02192

```
#authors      Soussi, T.; de Fromental, C.C.; Breugnot, C.; May, E.
#journal      Nucleic Acids Res. (1988) 16:11384
#title        Nucleotide sequence of a cDNA encoding the rat p53 nuclear
               oncoprotein
#cross-references MUID:89083585
#accession     S02192
#molecule_type mRNA
##residues     1-391 #label SOU
##cross-references EMBL:X13058; NID:g56828; PID:g56829
REFERENCE      S41149
#authors      Hulla, J.E.; Schneider, R.P.
#journal      Nucleic Acids Res. (1993) 21:713-717
#title        Structure of the rat p53 tumor suppressor gene.
#cross-references MUID:93181268
#accession     S41149
#status        preliminary; nucleic acid sequence not shown;
               translation not shown
#molecule_type DNA
##residues     1-173, 'W', 175-391 #label HUL
##cross-references EMBL:L07909
##note         the nucleotide sequence was submitted to the EMBL Data
               Library, December 1992
GENETICS       25/2; 32/3; 123/3; 185/1; 259/2; 305/1; 329/3; 365/2
#introns       25/2; 32/3; 123/3; 185/1; 259/2; 305/1; 329/3; 365/2
CLASSIFICATION #superfamily cellular tumor antigen p53
KEYWORDS       apoptosis; cell division control; DNA binding; homotetramer;
               nucleus; phosphoprotein; transcription regulation; tumor
               suppressor; zinc
FEATURE        174,177,236,240 #binding_site zinc (Cys, His, Cys, Cys) #status
               predicted\
390             #binding_site phosphoryl-RNA (Ser) (covalent) #status
               predicted
SUMMARY        #length 391 #molecular-weight 43451 #checksum 7105
               Query Match 100.0%; Score 63; DB 2; Length 391;
               Best Local Similarity 100.0%; Pred. No. 2.91e-03;
               Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 152 GTRVRAMAI 160
|||||||
QY 1 GTRVRAMAI 9

RESULT 3
ENTRY JC6193 #type complete
TITLE tumor suppressor p53 - rabbit
ALTERNATE_NAMES #formal_name Oryctolagus cuniculus #common_name domestic
ORGANISM rabbit
DATE 11-Apr-1997 #sequence_revision 09-May-1997 #text_change
17-Mar-1999
ACCESSIONS JC6193
REFERENCE JC6193
#authors Le Goas, F.; May, P.; Ronco, P.; de Fromental, C.C.
#journal Gene (1997) 185:169-173
#title cDNA cloning and immunological characterization of rabbit
               p53.
#cross-references MUID:97208869
#accession JC6193
#molecule_type mRNA
##residues 1-391 #label LFA
##cross-references EMBL:X90592; NID:gl532043; PID:e194962; PID:gl532044
GENETICS       p53
#gene        #superfamily cellular tumor antigen p53
CLASSIFICATION #superfamily cellular tumor antigen p53
KEYWORDS       tumor
SUMMARY        #length 391 #molecular-weight 43435 #checksum 4367
               Query Match 100.0%; Score 63; DB 2; Length 391;
               Best Local Similarity 100.0%; Pred. No. 2.91e-03;
               Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 151 GTRVRAMAI 159
|||||||
QY 1 GTRVRAMAI 9

RESULT 4
ENTRY JC6176 #type complete
TITLE tumor suppressor protein p53 - Chinese hamster
ALTERNATE_NAMES #formal_name Cricetus griseus #common_name Chinese hamster
ORGANISM hamster
DATE 11-Apr-1997 #sequence_revision 09-May-1997 #text_change
08-Sep-1997
ACCESSIONS JC6176
REFERENCE JC6176
#authors Lee, H.; Larner, J.M.; Hamlin, J.L.
#journal Gene (1997) 184:177-183
#title Cloning and characterization of Chinese hamster p53 cDNA.
#cross-references MUID:97183659
#contents     liver
#accession     JC6176
#molecule_type mRNA
##residues 1-393 #label LEE
##cross-references GB:U50395; NID:gl842229; PID:gl842230
COMMENT        This protein is a multimer, it plays the central role in a complex
               DNA damage-sensing network. It binds to replication factor and
               TATA-binding protein, and affects DNA replication, transcription,
               and recombination by protein/protein interactions.
GENETICS       p53
#gene        #superfamily cellular tumor antigen p53
CLASSIFICATION #superfamily cellular tumor antigen p53
KEYWORDS       liver; tumor
SUMMARY        #length 393 #molecular-weight 43362 #checksum 4043
               Query Match 100.0%; Score 63; DB 2; Length 393;
               Best Local Similarity 100.0%; Pred. No. 2.91e-03;
               Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 154 GTRVRAMAI 162
|||||||
QY 1 GTRVRAMAI 9

RESULT 5
ENTRY DNHU53 #type complete
TITLE cellular tumor antigen p53 - human
ALTERNATE_NAMES cellular phosphoprotein p53; oncoprotein p53; transformation
               suppressor p53; tumor suppressor p53
ORGANISM #formal_name Homo sapiens #common_name man
DATE 05-Oct-1988 #sequence_revision 18-Nov-1994 #text_change
26-Feb-1999
ACCESSIONS A25224; A43073; J0436; S40773; S42569; A22837; A55060;
               A25397; B25397; S42452; S42453; I38082; I38083; I38084;
               I38085; I38086; I38087; I38088; I38089; I38090; I38091;
               I38092; I38093; A44905; I58354; I78850; I52681; S60153
REFERENCE A25224
#authors Lamb, P.; Crawford, L.
#journal Mol. Cell. Biol. (1986) 6:1379-1385
#title Characterization of the human p53 gene.
#cross-references MUID:87064416
#accession A25224
#molecule_type DNA
##residues 1-393 #label LAM
##cross-references EMBL:X01405; GB:MI3121; GB:N00032; NID:gl89460;
               PID:g386994
REFERENCE J0436
#authors Buchman, V.L.; Chumakov, P.M.; Ninkina, N.N.; Samarina, O.P.;
               Georgiev, G.P.
#journal Gene (1988) 70:245-252
#title A variation in the structure of the protein-coding region of
               the human p53 gene.
#cross-references MUID:89108008
#accession A43073
#molecule_type DNA
##residues 1-393 #label BUC1
##cross-references EMBL:M22898; NID:gl89474
```

##note this 72-Arg allele appears to be about 5 times more frequent than the 72-Pro allele

##accession JT0436

##molecule\_type DNA

##residues 1-71,'P',73-393 ##label BUC2

##cross-references EMBL:M22898; NID:g189474; PID:g189476

##note this 72-Pro allele was found in both normal and malignant cell lines

REFERENCE S40773

##molecule\_type DNA

##residues 1-393 ##label CHU

##cross-references EMBL:X54156; NID:g35213; PID:g35214

REFERENCE S42669

##molecule\_type DNA

##residues 101-393 ##label MK11

##cross-references EMBL:X01405; NID:g35215; PID:g642241

REFERENCE A22837

##molecule\_type mRNA

##residues 1-192,'R',194-393 ##label F02

##cross-references EMBL:X60011; NID:g506434; PID:g506435

REFERENCE EMB0 J. (1984) 3:1257-3262

##molecule\_type mRNA

##residues 1-189,'LLSILSEWKEICVMSIMTETLFDIVWCPMSRLRLALT', 'VPPSTTTTCVTTPAWAA' ##label F01

##cross-references EMBL:X60010; NID:g506432; PID:g506433

##note deletion of a C nucleotide causes a frameshift at position 566

REFERENCE A22837

##molecule\_type mRNA

##residues 1-393 ##label ZAK

##cross-references EMBL:X02469; EMBL:M60950; NID:g35209; PID:g35210

REFERENCE A55060

##molecule\_type mRNA

##residues 1-71,'P',73-272,'H',274-393 ##label HAR

##cross-references GB:K03199; NID:g189478; PID:g189479

REFERENCE A93086

##molecule\_type mRNA

##residues 1-247,'Q',249-393 ##label F06

##cross-references EMBL:X60015; NID:g506442; PID:g506443

REFERENCE A93086

##molecule\_type mRNA

##residues 1-247,'Q',249-393 ##label F08

##cross-references EMBL:X60017; NID:g506446; PID:g506447

REFERENCE A93086

##molecule\_type mRNA

##residues 1-212,'Q',214-393 ##label F10

##cross-references EMBL:X60019; NID:g506450; PID:g506451

REFERENCE A93086

##molecule\_type mRNA

##residues 1-253,'D',255-393 ##label F11

##accession S42452

##molecule\_type mRNA; DNA

##residues 66-71,'P',73-79 ##label MK12

##experimental\_source clone lambda C113

##note 72-Cys was also found, and appears to represent a polymorphism

REFERENCE S42453

##molecule\_type mRNA; DNA

##residues 66-79 ##label MK13

##experimental\_source clone J6K

REFERENCE I38082

##molecule\_type mRNA

##residues 1-189,'LLSILSEWKEICVMSIMTETLFDIVWCPMSRLRLALT', 'VPPSTTTTCVTTPAWAA' ##label F01

##cross-references EMBL:X60010; NID:g506432; PID:g506433

##note deletion of a C nucleotide causes a frameshift at position 566

REFERENCE I38083

##molecule\_type mRNA

##residues 1-192,'R',194-393 ##label F02

##cross-references EMBL:X60011; NID:g506434; PID:g506435

REFERENCE I38084

##molecule\_type mRNA

##residues 1-393 ##label F03

##cross-references EMBL:X60012; NID:g506436; PID:g506437

REFERENCE I38085

##molecule\_type mRNA

##residues 1-245,'T',247-393 ##label F04

##cross-references EMBL:X60013; NID:g506438; PID:g506439

REFERENCE I38086

##molecule\_type mRNA

##residues 1-236,'I',238-393 ##label F05

##cross-references EMBL:X60014; NID:g506440; PID:g506441

REFERENCE I38087

##molecule\_type mRNA

##residues 1-247,'Q',249-393 ##label F06

##cross-references EMBL:X60015; NID:g506442; PID:g506443

REFERENCE I38088

##molecule\_type mRNA

##residues 1-71,'P',73-237,'Y',239-393 ##label F07

##cross-references EMBL:X60016; NID:g506444; PID:g506445

REFERENCE I38089

##molecule\_type mRNA

##residues 1-247,'Q',249-393 ##label F08

##cross-references EMBL:X60017; NID:g506446; PID:g506447

REFERENCE I38090

##molecule\_type mRNA

##residues 1-212,'Q',214-393 ##label F10

##cross-references EMBL:X60019; NID:g506450; PID:g506451

REFERENCE I38092

##molecule\_type mRNA

##residues 1-253,'D',255-393 ##label F11

```

REFERENCE
#authors Futreal, P.A.; Barrett, J.C.; Wiseman, R.W.
#journal Nucleic Acids Res. (1991) 19:6977
#title An Alu polymorphism intragenic to the TP53 gene.
#cross-references MUID:92107726
#accession I38093
#status translated from GB/EMBL/DBJ
#molecule_type DNA
#residues 1-393 ##label FUT
#cross-references EMBL:X54156; NID:g35213; PID:g35214
#authors Yamada, Y.; Yoshida, T.; Hayashi, K.; Sekiya, T.; Yokota, J.; Hirohashi, S.; Nakatani, K.; Nakano, H.; Sugimura, T.; Terada, M.
#journal Cancer Res. (1991) 51:5800-5805
#title p53 gene mutations in gastric cancer metastases and in gastric cancer cell lines derived from metastases.
#cross-references MUID:92034678
#accession A44905
...
Note: remainder of annotations omitted.
Query Match 100.0%; Score 63; DB 1; Length 393;
Best Local Similarity 100.0%; Pred. No. 2,91e-03;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 154 GTRVRAMAI 162
|||||||
QY 1 GTRVRAMAI 9

RESULT 6
ENTRY JH0633 #type complete
TITLE cellular tumor antigen p53 - golden hamster
ALTERNATE_NAMES tumor-suppressor protein p53
ORGANISM #formal_name Mesocricetus auratus #common_name golden hamster
DATE 17-Aug-1992 #sequence_revision 17-Aug-1992 #text_change 08-Sep-1997
ACCESSIONS JH0633
REFERENCE JH0633
#authors Legros, Y.; McIntyre, P.; Soussi, T.
#journal Gene (1992) 112:247-250
#title The cDNA cloning and immunological characterization of hamster p53.
#cross-references MUID:92210007
#accession JH0633
#molecule_type mRNA
#residues 1-396 ##label LEG
#cross-references GB:W5144; NID:g191414; PID:g191415
##experimental_source kidney, strain MPl

GENETICS
#gene p53
CLASSIFICATION #superfamily cellular tumor antigen p53
KEYWORDS apoptosis; cell division control; DNA binding; homotetramer; nucleus; phosphoprotein; transcription regulation; tumor suppressor; zinc

FEATURE
179,182,241,245 #binding_site zinc (Cys, His, Cys, Cys) #status predicted
395 #binding_site phosphoryl-RNA (Ser) (covalent) #status predicted
SUMMARY #length 396 #molecular_weight 43631 #checksum 6617

Query Match 100.0%; Score 63; DB 2; Length 396;
Best Local Similarity 100.0%; Pred. No. 2,91e-03;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 157 GTRVRAMAI 165
|||||||
QY 1 GTRVRAMAI 9

```



ALTERNATE\_NAMES oncoprotein p53  
 ORGANISM #formal\_name Mus musculus #common\_name house mouse  
 DATE 28-Aug-1985 #sequence\_revision 04-Oct-1996 #text\_change 12-Feb-1999

ACCESSIONS A22739; S06336; A02684; S38822; S40014; I48703  
 REFERENCE A22739  
 #authors Bizen, B.; Zakut-Houri, R.; Givol, D.; Oren, M.  
 #journal EMBO J. (1984) 3:2179-2183  
 #cross-references MUID:85027173  
 #accession A22739  
 ##molecule\_type DNA  
 ##residues 1-134, 'V', 136-390 #label B1E  
 ##cross-references GB:X00876; NID:9871420; PID:g871421; GB:X01237; GB:X01700; NID:g53575; PID:g53576

REFERENCE S06336  
 #authors Chumakov, P.M.  
 #journal Bioorg. Khim. (1987) 13:1691-1694  
 #title Primary structure of DNA complementary to murine oncoprotein p53 mRNA  
 #cross-references MUID:88221682  
 #accession S06336  
 ##status not compared with conceptual translation  
 ##molecule\_type mRNA  
 ##residues 1-134, 'V', 136-390 #label CHU

REFERENCE A02684  
 #authors Zakut-Houri, R.; Oren, M.; Bizen, B.; Lavie, V.; Hazum, S.; Givol, D.  
 #journal Nature (1983) 306:594-597  
 #title A single gene and a pseudogene for the cellular tumour antigen p53.  
 #cross-references MUID:84068204  
 #accession A02684  
 ##molecule\_type mRNA  
 ##residues 1-159, 'H', 161-167, 'G', 169-233, 'I', 235-390 #label ZAK  
 ##cross-references GB:X01237; GB:X01700; NID:g53575

REFERENCE S38822  
 #authors Arai, N.; Nomura, D.; Yokota, K.; Wolf, D.; Brill, E.; Shohat, O.; Rotter, V.  
 #journal Mol. Cell. Biol. (1986) 6:3232-3239  
 #title Immunologically distinct p53 molecules generated by alternative splicing.  
 #cross-references MUID:87064640  
 #accession S38822  
 ##status preliminary  
 ##molecule\_type mRNA  
 ##residues 1-390 #label ARA1  
 ##cross-references EMBL:M13872; NID:g200198; PID:g200199  
 #accession S38823  
 ##status preliminary  
 ##molecule\_type mRNA  
 ##residues 1-167, 'G', 169-233, 'I', 235-390 #label ARA2  
 ##cross-references EMBL:M13873

REFERENCE S40014  
 #authors Arai, N.; Nomura, D.; Yokota, K.; Wolf, D.; Brill, E.; Shohat, O.; Rotter, V.  
 #submission submitted to the EMBL Data Library, July 1988  
 #accession S40014  
 ##molecule\_type mRNA  
 ##residues 1-167, 'G', 169-390 #label ARA3  
 ##cross-references EMBL:M13873; NID:g200200; PID:g200201

REFERENCE I48703  
 #authors Jenkins, J.R.; Rudge, K.; Redmond, S.; Wade-Evans, A.  
 #journal Nucleic Acids Res. (1984) 12:5603-5626  
 #title Cloning and expression analysis of full length mouse cDNA sequences encoding the transformation associated protein p53  
 #cross-references MUID:84272240  
 #accession I48703  
 ##status preliminary; translated from GB/EMBL/DBJ  
 ##molecule\_type mRNA  
 ##residues 1-47, 'R', 49-78, 'OW', 82-390 #label RES  
 ##cross-references EMBL:X00741; NID:g53570; PID:g53571

COMMENT This DNA-binding protein plays an essential role in the regulation

of cell division, as it is required for the transition from phase G0 to G1 of the cell cycle.  
 The tetramer association region may exhibit a beta-turn, beta-sheet, beta-turn, alpha-helix motif.  
 CLASSIFICATION #superfamily cellular tumor antigen p53  
 KEYWORDS apoptosis; cell division control; DNA binding; homotetramer; phosphoprotein; transcription regulation; tumor suppressor; zinc

FEATURE 1-44 #domain transcription activation #status predicted

16-26 #label TRA  
 99-289 #region conserved region I  
 108-121 #domain DNA-binding core #status predicted #label DBC  
 114-139 #region L1 loop  
 160-192 #region conserved region II  
 168-178 #region L2 loop  
 231-252 #region conserved region III  
 233-248 #region conserved region IV  
 267-283 #region L3 loop  
 313-319 #region conserved region V  
 319-357 #region nuclear location signal  
 7,9,12,18,23,37 #region tetramer association  
 #binding\_site phosphate (Ser) (covalent) #status predicted  
 173,176,235,239 #binding\_site zinc (Cys, His, Cys, Cys) #status predicted  
 312 #binding\_site phosphate (Ser) (covalent) (by cdc2 kinase) #status predicted  
 389 #binding\_site phosphoryl-RNA (Ser) (covalent) #status predicted

SUMMARY #length 390 #molecular-weight 43458 #checksum 1260

Query Match 95.2%; Score 60; DB 1; Length 390;  
 Best Local Similarity 88.9%; Pred. No. 1.56e-02;  
 Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

DB 151 GSRVRAMAI 159

QY 1 GTRVRAMAI 9

RESULT 9 S06594 #type complete  
 ENTRY cellular tumor antigen p53 - green monkey  
 TITLE #formal\_name Cercopithecus aethiops #common\_name green monkey, grivet  
 ORGANISM 28-Feb-1990 #sequence\_revision 28-Feb-1990 #text\_change 08-Sep-1997  
 DATE S06594  
 ACCESSIONS S06594  
 REFERENCE S06594  
 #authors Rigaudy, P.; Eckhart, W.  
 #journal Nucleic Acids Res. (1989) 17:8375  
 #title Nucleotide sequence of a cDNA encoding the monkey cellular phosphoprotein p53.  
 #cross-references MUID:90045967  
 #accession S06594  
 ##molecule\_type mRNA  
 ##residues 1-393 #label RIG  
 ##cross-references EMBL:X15384; NID:g22795; PID:g22796

CLASSIFICATION #superfamily cellular tumor antigen p53  
 KEYWORDS apoptosis; cell division control; DNA binding; homotetramer; nucleus; phosphoprotein; transcription regulation; tumor suppressor; zinc

FEATURE 176,179,238,242 #binding\_site zinc (Cys, His, Cys, Cys) #status predicted  
 392 #binding\_site phosphoryl-RNA (Ser) (covalent) #status predicted

SUMMARY #length 393 #molecular-weight 43696 #checksum 4263

Query Match 95.2%; Score 60; DB 2; Length 393;  
 Best Local Similarity 88.9%; Pred. No. 1.56e-02;  
 Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

Db 154 GSRVRAMAI 162
QY 1 GTRVRAMAI 9

RESULT 10
ENTRY #type complete
TITLE sensory transduction histidine kinase - Methanobacterium
ORGANISM thermoautotrophicum (strain Delta H)
DATE #formal_name Methanobacterium thermoautotrophicum
05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change
B69137
A69000
Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.;
Dubois, J.; Aldredge, T.; Bashirzadeh, R.; Blakely, D.;
Cook, R.; Gilbert, K.; Harrison, D.; Hoang, L.; Keagle, P.;
Lumm, W.; Pothier, B.; Qiu, D.; Spadafora, R.; Vicaire, R.;
Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwani, N.; Caruso,
A.; Bush, D.; Safer, H.; Patwell, D.; Prabhakar, S.;
McDougall, S.; Shimer, G.; Goyal, A.; Pietrokovski, S.;
Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling,
J.; Reeve, J.N.
J. Bacteriol. (1997) 179:7135-7155
Complete genome sequence of Methanobacterium
thermoautotrophicum Delta H: functional analysis and
comparative genomics.
#cross-references NUID:98037514
#accession B69137
#status preliminary; nucleic acid sequence not shown;
translation not shown
#molecule_type DNA
#residues 1-564 #label MTH
#cross-references GB:AE000814; GB:AE000666; NID:g2621334; PID:g2621344
#experimental_source strain Delta H
GENETICS
#gene MTH292
#start_codon GTG
SUMMARY #length 564 #molecular-weight 62258 #checksum 5546
Query Match 85.7%; Score 54; DB 2; Length 564;
Best Local Similarity 87.5%; Pred. No. 3.96e-01;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 409 TRVRAMTI 416
QY 2 TRVRAMAI 9

RESULT 11
ENTRY #type complete
TITLE sensory transduction histidine kinase - Methanobacterium
ORGANISM thermoautotrophicum (strain Delta H)
DATE #formal_name Methanobacterium thermoautotrophicum
05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change
B69145
A69000
Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.;
Dubois, J.; Aldredge, T.; Bashirzadeh, R.; Blakely, D.;
Cook, R.; Gilbert, K.; Harrison, D.; Hoang, L.; Keagle, P.;
Lumm, W.; Pothier, B.; Qiu, D.; Spadafora, R.; Vicaire, R.;
Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwani, N.; Caruso,
A.; Bush, D.; Safer, H.; Patwell, D.; Prabhakar, S.;
McDougall, S.; Shimer, G.; Goyal, A.; Pietrokovski, S.;
Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling,
J.; Reeve, J.N.
J. Bacteriol. (1997) 179:7135-7155
Complete genome sequence of Methanobacterium
thermoautotrophicum Delta H: functional analysis and
comparative genomics.
#cross-references NUID:98037514
#accession B69145
#status preliminary; nucleic acid sequence not shown;
translation not shown
#molecule_type DNA
#residues 1-567 #label MTH
#cross-references GB:AE000821; GB:AE000666; NID:g2621414; PID:g2621415
#experimental_source strain Delta H
GENETICS
#gene MTH356
#start_codon GTG
SUMMARY #length 567 #molecular-weight 62431 #checksum 5446
Query Match 84.1%; Score 53; DB 2; Length 567;
Best Local Similarity 87.5%; Pred. No. 6.66e-01;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 415 SRVRAMAI 422
QY 2 TRVRAMAI 9

RESULT 12
ENTRY #type complete
TITLE DNA-directed RNA polymerase (EC 2.7.7.6) - Sulfolobus
ORGANISM acidocaldarius
DATE #formal_name Sulfolobus acidocaldarius
09-Dec-1993 #sequence_revision 10-Nov-1995 #text_change
13-Sep-1998
S42389; S38658
S42389
Langer, D.; Lottspeich, F.; Zillig, W.
Nucleic Acids Res. (1994) 22:694
A subunit of an archaeal DNA-dependent RNA polymerase
contains the S1 motif.
#cross-references NUID:94173739
#accession S42389
#status preliminary
#molecule_type DNA
#residues 1-248 #label LA2
#cross-references EMBL:X75411; NID:g415998; PID:g415999
CLASSIFICATION superfamily DNA-directed RNA polymerase subunit E
KEYWORDS nucleotidyltransferase
SUMMARY #length 248 #molecular-weight 27632 #checksum 1833
Query Match 82.5%; Score 52; DB 2; Length 248;
Best Local Similarity 77.8%; Pred. No. 1.11e+00;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 139 GDRVRAMII 147
QY 1 GTRVRAMAI 9

RESULT 13
ENTRY #type complete
TITLE contractile tail sheath protein - Pseudomonas aeruginosa
ORGANISM phase PS17
DATE #formal_name Pseudomonas aeruginosa phage PS17
15-Aug-1996 #sequence_revision 18-Oct-1996 #text_change
10-Sep-1997
JC4865
JC4865
Sasaki, T.; Shinomiya, T.; Kumazaki, T.; Mohri, N.; Ishii,
S.; Ariaka, F.
submitted to JIPID, September 1996
#description Nucleotide sequences of the contractile tail sheath and tube
genes of bacteriophage PS17 and amino acid sequences of
their products.
#accession JC4865
#molecule_type DNA
#residues 1-386 #label SAS
#cross-references DDBJ:D26449; NID:g452162; PID:d1006009; PID:g514392
SUMMARY #length 386 #molecular-weight 41369 #checksum 2919

```

```

Query Match      79.4%; Score 50; DB 2; Length 386;
Best Local Similarity 75.0%; Pred. No. 3.05e+00;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 282 TRVRTMDI 289
|||||
QY 2 TRVRWAI 9

RESULT 14
ENTRY JC5191 #type complete
TITLE tail sheath protein - Pseudomonas aeruginosa phage PS17
ORGANISM #formal_name Pseudomonas aeruginosa phage PS17
DATE 20-Feb-1997 #sequence_revision 27-Feb-1997 #text_change 13-Jun-1997
ACCESSIONS JC5191
REFERENCE JC5191
#authors Sasaki, T.; Shinomiya, T.; Kumazaki, T.; Mohri, N.; Ishii, S.; Arisaka, F.
#journal Res. Commun. Biochem. Cell Mol. Biol. (1997) 1:93-107
#title Nucleotide sequences of the contractile tail sheath and tube genes of bacteriophage PS17 and amino acid sequences of their products.
#accession JC5191
#molecule_type DNA
#residues 1-386 #label SAS
#cross-references DBJ:D36449; NID:g452162; PID:g514392
GENETICS FI
#gene tail protein
KEYWORDS #product tail sheath protein #status predicted #label
FEATURE 2-386 WAT
SUMMARY #length 386 #molecular-weight 41369 #checksum 2919

Query Match      79.4%; Score 50; DB 2; Length 386;
Best Local Similarity 75.0%; Pred. No. 3.05e+00;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 282 TRVRTMDI 289
|||||
QY 2 TRVRWAI 9

RESULT 15
ENTRY XUEC #type complete
TITLE acetyl-CoA C-acyltransferase (EC 2.3.1.16) - Escherichia coli (strain K-12)
ALTERNATE_NAMES 3-ketoacyl-CoA thiolase; beta-ketothiolase; degradative thiolase; fatty acid beta oxidation multienzyme complex
ORGANISM #formal_name Escherichia coli
DATE 30-Jun-1991 #sequence_revision 10-Oct-1997 #text_change 20-Mar-1998
ACCESSIONS F65189; JQ0655; A35436; S30736; A40816
REFERENCE A64720
#authors Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.; Rode, C.K.; Mayhew, G.E.; Gregor, J.; Davis, N.W.; Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao, Y.
#journal Science (1997) 277:1453-1462
#title The complete genome sequence of Escherichia coli K-12.
#cross-references MUID:97426617
#accession F65189
#status preliminary; nucleic acid sequence not shown; translation not shown
#molecule_type DNA
#residues 1-387 #label BLAT
#cross-references GB:AE000460; GB:U00096; NID:g2367315; PID:g2367316; UWG:b3845
#experimental_source strain K-12, substrain MG1655

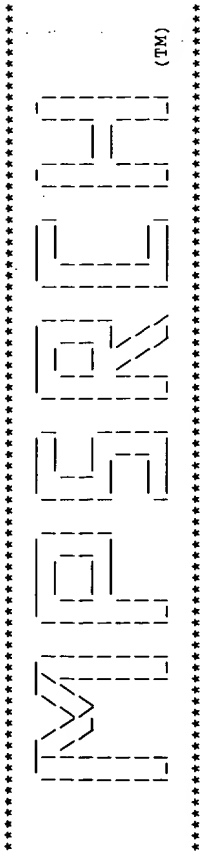
REFERENCE JVO108
#authors Dirusso, C.C.
#journal J. Bacteriol. (1990) 172:6459-6468
#title Primary sequence of the Escherichia coli fadBA operon, encoding the fatty acid-oxidizing multienzyme complex, indicates a high degree of homology to eucaryotic enzymes.
#cross-references MUID:91035260
#accession JVO109
#molecule_type DNA
#residues 1-36,'S',38-387 #label DIR
#cross-references GB:M59368; GB:M36149; NID:g145899; PID:g145901
REFERENCE JQ0654
#authors Nakanigashi, K.; Inokuchi, H.
#journal Nucleic Acids Res. (1990) 18:4937
#title Nucleotide sequence of the fadA and fadB genes from Escherichia coli.
#cross-references MUID:90370500
#accession JQ0655
#molecule_type DNA
#residues 1-118,'G',120-387 #label NAK
#cross-references EMBL:X52837
REFERENCE A35436
#authors Yang, S.Y.; Yang, X.Y.H.; Healy-Louie, G.; Schulz, H.; Eizinga, M.
#journal J. Biol. Chem. (1990) 265:10424-10429
#title Nucleotide sequence of the fadA gene. Primary structure of 3-ketoacyl-coenzyme A thiolase from Escherichia coli and the structural organization of the fadAB operon.
#cross-references MUID:90285166
#accession A35436
#molecule_type DNA
#residues 1-36,'S',38-370,'DG',373,'VS',375-387 #label YAN
#cross-references EMBL:J05498
REFERENCE S30660
#authors Daniels, D.L.; Plunkett III, G.; Burland, V.; Blattner, F.R.
#journal Science (1992) 257:771-778
#title Analysis of the Escherichia coli genome: DNA sequence of the region from 84.5 to 86.5 minutes.
#cross-references MUID:92358234
#accession S30736
#molecule_type DNA
#residues 1-48,'X',50-81,'X',83-170,'XX',173-339,'X',341-387 #label DAN
#cross-references EMBL:M87049
REFERENCE A40816
#authors Yang, S.Y.; Yang, X.Y.H.; Healy-Louie, G.; Schulz, H.; Eizinga, M.
#journal J. Biol. Chem. (1991) 266:16255
#cross-references MUID:91340783
#contents erratum
#accession A40816
#status preliminary
#molecule_type DNA
#residues 368-377 #label YA2
GENETICS
#gene fadA
#map_position 87 min
FUNCTION catalyzes the transfer of the acyl group from acyl-CoA to acetyl-CoA to form 3-oxoacyl-CoA, the last step of fatty acid beta-oxidation
#description acetyl-CoA to form 3-oxoacyl-CoA, the last step of fatty acid beta-oxidation
#pathway the E. coli enzyme is the beta chain of the fatty acid complex
#note #superfamily acetyl-CoA acetyltransferase
CLASSIFICATION #superfamily acetyl-CoA acetyltransferase
KEYWORDS acyltransferase; fatty acid beta-oxidation
FEATURE 91
SUMMARY #active_site Cys #status predicted
#length 387 #molecular-weight 40890 #checksum 7496

Query Match      77.8%; Score 49; DB 1; Length 387;
Best Local Similarity 62.5%; Pred. No. 5.00e+00;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

```

Db 265 ARVRSMAV 272  
:|||||:  
QY 2 TRVRAMAI 9

Search completed: Sat Apr 15 00:36:38 2000  
Job time : 19 secs.



\*\*\*\*\*  
Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd  
\*\*\*\*\*  
MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Sat Apr 15 00:36:55 2000; MasPar time 3.09 Seconds  
Tabular output not generated. 86.958 Million cell updates/sec  
\*\*\*\*\*

Title: >US-08-452-843-16  
Description: (1-9) from US08452843.ppe  
Perfect Score: 63  
Sequence: 1 GTRVRAMAI 9  
Scoring table: PAM 150  
Gap 15  
Searched: 82229 seqs, 29864866 residues  
Post-processing: Minimum Match 0%  
Listing first 45 summaries  
Database: swiss-prot38  
1:swissprot  
Statistics: Mean 23.405; Variance 23.631; scale 0.990

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES				Pred. No.	
Result No.	Score	Query Match	Description	ID	
1	63	100.0	CELLULAR TUMOR ANTIGEN	P53_EQUAS	3.34e-04
2	63	100.0	CELLULAR TUMOR ANTIGEN	P53_HORSE	3.34e-04
3	63	100.0	CELLULAR TUMOR ANTIGEN	P53_SHEEP	3.34e-04
4	63	100.0	CELLULAR TUMOR ANTIGEN	P53_SHEEP	3.34e-04
5	63	100.0	CELLULAR TUMOR ANTIGEN	P53_BOVIN	3.34e-04
6	63	100.0	CELLULAR TUMOR ANTIGEN	P53_RAT	3.34e-04
7	63	100.0	CELLULAR TUMOR ANTIGEN	P53_RABIT	3.34e-04
8	63	100.0	CELLULAR TUMOR ANTIGEN	P53_HUMAN	3.34e-04
9	63	100.0	CELLULAR TUMOR ANTIGEN	P53_CRIGR	3.34e-04
10	63	100.0	CELLULAR TUMOR ANTIGEN	P53_MESAU	3.34e-04
11	60	95.2	CELLULAR TUMOR ANTIGEN	P53_MOUSE	2.27e-03
12	60	95.2	CELLULAR TUMOR ANTIGEN	P53_CERAE	2.27e-03
13	60	95.2	CELLULAR TUMOR ANTIGEN	P53_MACMU	2.27e-03
14	60	95.2	CELLULAR TUMOR ANTIGEN	P53_MACFA	2.27e-03
15	52	82.5	DNA-DIRECTED RNA POLYM	RRA_PINTH	2.86e-01
16	49	77.8	CHLOROPLAST 30S RIBOSO	P53_FELCA	1.56e+00
17	49	77.8	CELLULAR TUMOR ANTIGEN	P53_FELCA	1.56e+00
18	49	77.8	3-KETOACYL-COA THIOLAS	4-ALPHA-GLUCANOTRANSE	4.63e+00
19	47	74.6	HYPOHETICAL 55.1 KD P	MAIQ_THEAQ	4.63e+00
20	47	74.6	HYPOHETICAL 55.1 KD P	Y233_MYCTU	4.63e+00
21	47	74.6	DNA LIGASE (EC 6.5.1.1	DNLL_METJA	4.63e+00
22	46	73.0	TRNA (GUANOSINE-2'-O-)	TRNH_ECOLI	7.88e+00
23	46	73.0	BIPHENYL DIOXYGENASE S	BPHG_BURCE	7.88e+00

KW Nuclear protein; Phosphorylation; Apoptosis.  
FT NON\_TER 1  
FT DOMAIN 187 199 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
FT NON\_TER 207 207  
SQ SEQUENCE 207 AA; 23428 MW; 0FBAE9C1 CRC32;  
Query Match 100.0%; Score 63; DB 1; Length 207;  
Best Local Similarity 100.0%; Pred. No. 3.34e-04;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Db 29 GTRVRAMAI 37  
QY 1 GTRVRAMAI 9  
RESULT 2  
ID P53\_HORSE STANDARD; PRT; 280 AA.  
AC P79892; Q29481;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).  
GN TP53 OR P53.  
OS Equus caballus (Horse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Perissodactyla; Equidae; Equus.  
RN [1]  
RP SEQUENCE OF 1-263 FROM N.A.  
RC TISSUE=SPLEEN;  
RX MEDLINE; 97070350.  
RA PAZZI K.A., KRAEGL S.A., GRIFFEY S.M., THEON A.P., MADEWELL B.R.;  
RT "Analysis of the equine tumor suppressor gene p53 in the normal horse  
and in eight cutaneous squamous cell carcinomas.";  
RL Cancer Lett. 107:125-130(1996).  
RN [2]  
RP SEQUENCE OF 76-280 FROM N.A.  
RX MEDLINE; 96293865.  
RA NASIR L., REID S.W.;  
RT "Nucleotide sequence of exons 5 to 9 of the p53 tumour-suppressor  
gene of the horse (Equus caballus).";  
RL DNA Seq. 6:185-187(1996).  
CC -1- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES  
GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL  
CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN  
TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION, IT IS A  
TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION  
BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF  
THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.  
APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF  
BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2  
EXPRESSION (BY SIMILARITY).  
CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
CC -1- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY  
OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED  
IN MANY TYPES OF CANCER.  
CC -1- SIMILARITY: BELONGS TO THE P53 FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL Outstation  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; S83123; AAB46899.1; -;  
CC EMBL; U37120; AAB18936.1; -;  
CC HSSP; P04637; 1SAH.  
CC PROSITE; PS00348; P53; 1.  
CC PFAM; PF00870; P53; 1.  
CC Anti-oncogene; DNA-binding; Transcription regulation; Activator;  
KW Nuclear protein; Phosphorylation; Apoptosis.  
FT NON\_TER 1

FT DOMAIN 262 274 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
FT CONFLICT 79 79 T -> A (IN REF. 2).  
FT CONFLICT 83 83 L -> M (IN REF. 2).  
FT CONFLICT 111 111 A -> V (IN REF. 2).  
FT CONFLICT 138 138 G -> A (IN REF. 2).  
FT NON\_TER 280 280  
SQ SEQUENCE 280 AA; 30985 MW; B494F872 CRC32;  
Query Match 100.0%; Score 63; DB 1; Length 280;  
Best Local Similarity 100.0%; Pred. No. 3.34e-04;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Db 104 GTRVRAMAI 112  
QY 1 GTRVRAMAI 9  
RESULT 3  
ID P53\_SPEBE STANDARD; PRT; 314 AA.  
AC Q64562;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).  
GN TP53.  
OS Sperophilus beecheyi (Beechey ground squirrel).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Rodentia; Sciurognathi; Scluridae; Sciurinae; Spermophilus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=THYMUS;  
RX MEDLINE; 95007566.  
RA RIVKINA M.B., CULLEN J.M., ROBINSON W.S., MARION P.L.;  
RT "State of the p53 gene in hepatocellular carcinomas of ground  
squirrels and woodchucks with past and ongoing infection with  
hepatidnaviruses.";  
RL Cancer Res. 54:5430-5437(1994).  
CC -1- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES  
GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL  
CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN  
TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION, IT IS A  
TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION  
BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF  
THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.  
APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF  
BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2  
EXPRESSION.  
CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
CC -1- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY  
OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED  
IN MANY TYPES OF CANCER.  
CC -1- SIMILARITY: BELONGS TO THE P53 FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL Outstation  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; U43902; AAA85628.1; -;  
CC HSSP; P04637; 1YCS.  
CC PROSITE; PS00348; P53; 1.  
CC PFAM; PF00870; P53; 1.  
CC Anti-oncogene; DNA-binding; Transcription regulation; Activator;  
KW Nuclear protein; Phosphorylation.  
FT NON\_TER 1  
FT DOMAIN 289 301  
FT NON\_TER 314 314  
SQ SEQUENCE 314 AA; 34618 MW; D07F433B CRC32;  
Query Match 100.0%; Score 63; DB 1; Length 314;

Best Local Similarity 100.0%; Pred. No. 3.34e-04; Mismatches 0; Indels 0; Gaps 0;  
Matches 9; Conservative 0;

Db 132 GTRVRAMAI 140  
|||||||  
QY 1 GTRVRAMAI 9

RESULT 4  
ID P53\_SHEEP STANDARD; PRT; 382 AA.  
AC P51664;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE CELLULAR TUMOR ANTIGEN P53.  
GN TP53.  
OS Ovis aries (Sheep).  
OC Eukaryota; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
OC Caprinae; Ovis.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-BLOOD;  
RX MEDLINE; 95352828.  
RA DEQUIEDT F., KETTMANN R., BURNY A., WILLEMS L.;  
RT "Nucleotide sequence of the ovine p53 tumor-suppressor cDNA and its  
RT genomic organization.";  
RL DNA Seq. 5:255-259(1995).  
CC -!- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES  
CC GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL  
CC CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN  
CC TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION, IT IS A  
CC TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION  
CC BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF  
CC THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.  
CC APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF  
CC BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2  
CC EXPRESSION.  
CC -!- SUBCELLULAR LOCATION: NUCLEAR.  
CC -!- SIMILARITY: BELONGS TO THE P53 FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; X81705; CAA57349.1; -  
CC HSSP; P04637; 1PPT.  
CC PROSITE; PS00348; P53; 1.  
CC PFAM; PF00870; P53; 1.  
CC Anti-oncogene; DNA-binding; Transcription regulation; Activator;  
CC Nuclear protein; Phosphorylation; Apoptosis.  
CC DOMAIN 1 66 ASP/GLU-RICH (ACIDIC).  
CC FT DOMAIN 300 312 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
CC FT MOD\_RES 381 381 PHOSPHORYLATION (BY SIMILARITY).  
CC SEQUENCE 382 AA; 42809 MW; 0CB99A00 CRC32;

Query Match 100.0%; Score 63; DB 1; Length 382;  
Best Local Similarity 100.0%; Pred. No. 3.34e-04;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 142 GTRVRAMAI 150  
|||||||  
QY 1 GTRVRAMAI 9

RESULT 5  
ID P53\_BOVIN STANDARD; PRT; 386 AA.  
AC Q29628;  
DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 15-JUL-1998 (Rel. 36, Last annotation update)  
DE CELLULAR TUMOR ANTIGEN P53.  
GN TP53.  
OS Bos taurus (Bovine), and Bos indicus (Zebu).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
OC Bovinae; Bos.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC SPECIES-BOVIN; TISSUE-LIVER;  
RX MEDLINE; 95352829.  
RA DEQUIEDT F., KETTMANN R., BURNY A., WILLEMS L.;  
RT "Nucleotide sequence of the bovine p53 tumor-suppressor cDNA.";  
RL DNA Seq. 5:261-264(1995).  
RN [2]  
RP SEQUENCE OF 13-386 FROM N.A.  
RC SPECIES-BOVIN; STRAIN-HOLSTEIN; TISSUE-THYMUS;  
RX MEDLINE; 96401400.  
RA KOMORI H., ISHIGURO N., HORIUCHI M., SHINAGAWA M., AIDA Y.;  
RT "Predominant p53 mutations in enzootic bovine leukemic cell lines.";  
RL Vet. Immunol. Immunopathol. 52:53-63(1996).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC SPECIES-B.INDICUS; STRAIN-BORAN; TISSUE-BLOOD;  
RA BISHOP R.R.P., GOBRIGHT E.E.I.;  
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES  
CC GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL  
CC CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN  
CC TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION, IT IS A  
CC TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION  
CC BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF  
CC THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.  
CC APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF  
CC BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2  
CC EXPRESSION.  
CC -!- SUBCELLULAR LOCATION: NUCLEAR.  
CC -!- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY  
CC OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED  
CC IN MANY TYPES OF CANCER.  
CC -!- SIMILARITY: BELONGS TO THE P53 FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; X81704; CAA57348.1; -  
CC HSSP; P04637; 1YCR.  
CC PROSITE; PS00348; P53; 1.  
CC PFAM; PF00870; P53; 1.  
CC Anti-oncogene; DNA-binding; Transcription regulation; Activator;  
CC Nuclear protein; Phosphorylation; Apoptosis.  
CC DOMAIN 1 59 ASP/GLU-RICH (ACIDIC).  
CC FT DOMAIN 304 316 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
CC FT MOD\_RES 385 385 PHOSPHORYLATION (BY SIMILARITY).  
CC FT CONFLICT 380 380 R -> T (IN REF. 2).  
CC SEQUENCE 386 AA; 43255 MW; 0322BF3D CRC32;

Query Match 100.0%; Score 63; DB 1; Length 386;  
Best Local Similarity 100.0%; Pred. No. 3.34e-04;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 146 GTRVRAMAI 154  
|||||||  
QY 1 GTRVRAMAI 9

```

RESULT 6
ID P53_RAT STANDARD; PRT; 391 AA.
AC P10361; O09168;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE CELLULAR TUMOR ANTIGEN P53.
GN TP53 OR P53.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
[1]
RN SEQUENCE FROM N.A.
RA SOUSSI T.;
RX MEDLINE; 89083585.
RT "Nucleotide sequence of a cDNA encoding the rat p53 nuclear
  oncoprotein.";
RL Nucleic Acids Res. 16:11384-11384(1988).
[2]
RN SEQUENCE FROM N.A.
RX MEDLINE; 93181268.
RA HULLA J.E., SCHNEIDER R.P.;
RT "Structure of the rat p53 tumor suppressor gene.";
RL Nucleic Acids Res. 21:713-717(1993).
[3]
RN SEQUENCE FROM N.A.
RC STRAIN-SPRAGUE-DAWLEY;
RA MATHUPALA S.P.;
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES
  GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL
  CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN
  TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION, IT IS A
  TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION
  BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF
  THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.
  APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF
  BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2
  EXPRESSION.
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY
  OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED
  IN MANY TYPES OF CANCER.
CC -!- SIMILARITY: BELONGS TO THE P53 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X13058; CAA31457.1; -
DR EMBL; L07910; AAA41788.1; -
DR EMBL; L07904; AAA41788.1; JOINED.
DR EMBL; L07905; AAA41788.1; JOINED.
DR EMBL; L07906; AAA41788.1; JOINED.
DR EMBL; L07907; AAA41788.1; JOINED.
DR EMBL; L07908; AAA41788.1; JOINED.
DR EMBL; L07909; AAA41788.1; JOINED.
DR EMBL; U90328; AAB80959.1; -
DR PIR; S02192; S02192.
DR HSP; P04637; 1PT.
DR PROSITE; PS00348; P53; 1.
DR PFAM; PF00870; P53; 1.
KW Anti-oncogene; DNA-binding; Transcription regulation; Activator;
  Nuclear protein; Phosphorylation; Apoptosis.
FT DOMAIN 1 76 ASP/GLU-RICH (ACIDIC).
FT DOMAIN 77 151 HYDROPHOBIC.
FT DOMAIN 277 391 HIGHLY BASIC AND MAY BE INVOLVED IN
  INTERACTION WITH DNA.
FT

```

```

FT DOMAIN 309 321 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT FT MOD_RES 390 390 PHOSPHORYLATION (BY SIMILARITY).
FT VARIANT 103 103 G -> S.
FT VARIANT 256 256 E -> G.
FT CONFLICT 174 174 C -> W (IN REF. 2).
SQ SEQUENCE 391 AA; 43451 MW; E0114C18 CRC32;

Query Match 100.0%; Score 63; DB 1; Length 391;
Best Local Similarity 100.0%; Pred. No. 3.34e-04;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 152 GTRVRAMAI 160
|||||||
QY 1 GTRVRAMAI 9

RESULT 7
ID P53_RABIT STANDARD; PRT; 391 AA.
AC Q95330;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE CELLULAR TUMOR ANTIGEN P53.
GN TP53.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Lagomorpha; Leporidae; Oryctolagus.
[1]
RN SEQUENCE FROM N.A.
RC STRAIN-NEW ZEALAND;
RX MEDLINE; 97208869.
RA LE GOAS F., MAY P., RONCO P., CARON DE FROMENTEL C.;
RT "CDNA cloning and immunological characterization of rabbit p53.";
RL Gene 185:169-173(1997).
CC -!- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES
  GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL
  CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN
  TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION, IT IS A
  TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION
  BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF
  THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.
  APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF
  BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2
  EXPRESSION (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY
  OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED
  IN MANY TYPES OF CANCER.
CC -!- SIMILARITY: BELONGS TO THE P53 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X90592; CAA62216.1; -
DR HSP; P04637; 1YCR.
DR PROSITE; PS00348; P53; 1.
DR PFAM; PF00870; P53; 1.
KW Anti-oncogene; DNA-binding; Transcription regulation; Activator;
  Nuclear protein; Phosphorylation; Apoptosis.
FT DOMAIN 1 70 ASP/GLU-RICH (ACIDIC).
FT DOMAIN 308 321 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT MOD_RES 390 390 PHOSPHORYLATION (BY SIMILARITY).
SQ SEQUENCE 391 AA; 43435 MW; 30A36172 CRC32;

Query Match 100.0%; Score 63; DB 1; Length 391;
Best Local Similarity 100.0%; Pred. No. 3.34e-04;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```



Db 151 GTRVRAMAI 159  
QY 1 GTRVRAMAI 9  
|||||  
RESULT 8  
ID P53\_HUMAN STANDARD; PRT; 393 AA.  
AC P04637;  
DT 13-AUG-1987 (Rel. 05, Created)  
DT 01-MAR-1989 (Rel. 10, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE CELLULAR TUMOR ANTIGEN P53 (PHOSPHOPROTEIN P53).  
GN TP53.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 85230577.  
RA ZAKUT-HOURI R., BIENZ-TADMOR B., GIVOL D., OREN M.;  
RT "Human p53 cellular tumor antigen: cDNA sequence and expression in  
RT COS cells.";  
RL EMBO J. 4:1251-1255(1985).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 87064416.  
RA LAMB P., CRAWFORD L.;  
RT "Characterization of the human p53 gene.";  
RL Mol. Cell. Biol. 6:1379-1385(1986).  
RN [3]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 85267676.  
RA HARLOW E., WILLIAMSON N.M., RALSTON R., HELFMAN D.M., ADAMS T.E.;  
RT "Molecular cloning and in vitro expression of a cDNA clone for human  
RT cellular tumor antigen p53.";  
RL Mol. Cell. Biol. 5:1601-1610(1985).  
RN [4]  
RP TRANSFORMED HYBRIDOMA SV-80 CELL LINE, SEQUENCE FROM N.A.  
RX MEDLINE; 87089826.  
RA HARRIS N., BRILL E., SHOHAT O., PROKOCIMER M., WOLF D., ARAI N.,  
RA ROTTER V.;  
RT "Molecular basis for heterogeneity of the human p53 protein.";  
RL Mol. Cell. Biol. 6:4650-4656(1986).  
RN [5]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 89108008.  
RA BUCHANAN V.L., CHURKOV P.M., NINKINA N.N., SAMARINA O.P.,  
RA GEORGIEV G.P.;  
RT "A variation in the structure of the protein-coding region of the  
RT human p53 gene.";  
RL Gene 70:245-252(1988).  
RN [6]  
RP SEQUENCE OF 101-393 FROM N.A.  
RX MEDLINE; 85126934.  
RA MATLASHESKI G., LAMB P., PIM D., PEACOCK J., CRAWFORD L.,  
RA BENCHIMOL S.;  
RT "Isolation and characterization of a human p53 cDNA clone: expression  
RT of the human p53 gene.";  
RL EMBO J. 3:3257-3262(1984).  
RN [7]  
RP NUCLEAR LOCALIZATION SIGNAL.  
RX MEDLINE; 90191730.  
RA ADDISON C., JENKINS J.R., STURZBECHER H.-W.;  
RT "The p53 nuclear localisation signal is structurally linked to a  
RT p34cdc2 kinase motif.";  
RL Oncogene 5:423-426(1990).  
RN [8]  
RP PHOSPHORYLATION BY P60/CDC2 AND CYCLIN B/CDC2.  
RX MEDLINE; 90280456.  
RA BISCHOFF J.R., FRIEDMAN P.N., MARSHAK D.R., PRIVES C., BEACH D.;  
RT "Human p53 is phosphorylated by p60-cdc2 and cyclin B-cdc2.";  
RL Proc. Natl. Acad. Sci. U.S.A. 87:4766-4770(1990).  
RN [9]  
RP DEPHOSPHORYLATION BY PP2A.  
RX MEDLINE; 91172186.  
RA SCHEIDTMANN K.H., MUMBY M.C., RUNDELL K., WALTER G.;  
RT "Dephosphorylation of simian virus 40 large-T antigen and p53 protein  
RT by protein phosphatase 2A: inhibition by small-t antigen.";  
RL Mol. Cell. Biol. 11:1996-2003(1991).  
RN [10]  
RP STRUCTURE BY NMR OF 319-360.  
RX MEDLINE; 94294808.  
RA CLORE G.M., OMICHINSKI J.G., SAKAGUCHI K., ZAMBRANO N., SAKAMOTO H.,  
RA APPELLA E., GRONENBORN A.M.;  
RT "High-resolution structure of the oligomerization domain of p53 by  
RT multidimensional NMR.";  
RL Science 265:386-391(1994).  
RN [11]  
RP STRUCTURE BY NMR OF 325-355.  
RX MEDLINE; 95292092.  
RA LEE W., HARVEY T.S., YIN Y., YAU P., LITCHFIELD D., ARROWSMITH C.H.;  
RT "Solution structure of the tetrameric minimum transforming domain of  
RT p53.";  
RL Nat. Struct. Biol. 1:877-890(1994).  
RN [12]  
RP STRUCTURE BY NMR OF 326-354.  
RX MEDLINE; 98026899.  
RA MCCOY M., STAVRIDIS E.S., WATERMAN J.L., WIECZOREK A.M., OPELLA S.J.,  
RA HALAZONETIS T.D.;  
RT "Hydrophobic side-chain size is a determinant of the  
RT three-dimensional structure of the p53 oligomerization domain.";  
RL EMBO J. 16:6230-6236(1997).  
RN [13]  
RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 94-289.  
RX MEDLINE; 94294806.  
RA CHO Y., GORINA S., JEFFREY P.D., PAVLETICH N.P.;  
RT "Crystal structure of a p53 tumor suppressor-DNA complex:  
RT understanding tumorigenic mutations.";  
RL Science 265:346-355(1994).  
RN [14]  
RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS) OF 13-29 IN COMPLEX WITH MDM2.  
RX MEDLINE; 97081050.  
RA KUSIE P.H., GORINA S., MARECHAL V., ELENAAS B., MOREAU J.,  
RA LEVINE A.J., PAVLETICH N.P.;  
RT "Structure of the MDM2 oncoprotein bound to the p53 tumor suppressor  
RT transactivation domain.";  
RL Science 274:948-953(1996).  
RN [15]  
RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 97-287 IN COMPLEX WITH 53BP2.  
RX MEDLINE; 97035414.  
RA GORINA S., PAVLETICH N.P.;  
RT "Structure of the p53 tumor suppressor bound to the ankyrin and SH3  
RT domains of 53BP2.";  
RL Science 274:1001-1005(1996).  
RN [16]  
RP REVIEW.  
RX MEDLINE; 94090335.  
RA HARRIS C.C.;  
RT "p53: at the crossroads of molecular carcinogenesis and risk  
RT assessment.";  
RL Science 262:1980-1981(1993).  
RN [17]  
RP REVIEW ON VARIANTS.  
RX MEDLINE; 91289156.  
RA HOOLSTEIN M., SIDRANSKY D., VOGELSTEIN B., HARRIS C.C.;  
RT "p53 mutations in human cancers.";  
RL Science 253:49-53(1991).  
RN [18]  
RP REVIEW ON VARIANTS.  
RX MEDLINE; 96271983.  
RA DE VRIES E.M.G., RICKE D.O., DE VRIES T.N., HARTMANN A., BLASZYK H.,  
RA LIAO D., SOUSI T., KOVACH J.S., SOMMER S.S.;  
RT "Database of mutations in the p53 and APC tumor suppressor genes  
RT designed to facilitate molecular epidemiological analyses.";  
RL Hum. Mutat. 7:202-213(1996).  
RN [19]

RP VARIANT ARG-72.  
RX MEDLINE; 91153807.  
RA OLSCHWANG S., LAURENT-PUIG P., VASSAL A., SALMON R.-J., THOMAS G.;  
RT "Characterization of a frequent polymorphism in the coding sequence  
of the TP53 gene in colonic cancer patients and a control  
population.";  
RL Hum. Genet. 86:369-370(1991).  
[20]  
RP VARIANT LFS THR-133.  
RX MEDLINE; 92034774.  
RA LAW J.C., STRONG L.C., CHIDAMBARAM A., FERRELL R.E.;  
RT "A germ line mutation in exon 5 of the p53 gene in an extended cancer  
family.";  
RL Cancer Res. 51:6385-6387(1991).  
[21]  
RP VARIANTS LFS CVS-245; TRP-248; PRO-252 AND LYS-258.  
RX MEDLINE; 91057657.  
RA MALKIN D., LI F.P., STRONG L.C., FRAUMENI J.F. JR., NELSON C.E.,  
KIM D.H., KASSEL J., GRYKA M.A., BISCHOFF F.Z., TAINSKY M.A.,  
FRIEND S.H.;  
RT "Germ line p53 mutations in a familial syndrome of breast cancer,  
sarcomas, and other neoplasms.";  
RL Science 250:1233-1238(1990).  
[22]  
RP VARIANT LFS ASP-245.  
RX MEDLINE; 91080929.  
RA SRIVASTAVA S., ZOU Z., PIROLLO K., BLATTNER W., CHANG E.H.;  
RT "Germ-line transmission of a mutated p53 gene in a cancer-prone  
family with Li-Fraumeni syndrome.";  
RL Nature 348:747-749(1990).  
[23]  
RP VARIANT LFS LEU-272.  
RX MEDLINE; 92147883.  
RA FELIX C.A., NAU M.M., TAKAHASHI T., MITSUDOMI T., CHIBA I.,  
RA POPLACK D.G., REAMAN G.H., COLE D.E., LETTERIO J.J., WHANG-PENG J.,  
KNOTSEN T., MINNA J.D.;  
RT "Hereditary and acquired p53 gene mutations in childhood acute  
lymphoblastic leukemia.";  
RL J. Clin. Invest. 89:640-647(1992).  
[24]  
RP VARIANTS LFS HIS-273 AND VAL-325.  
RX MEDLINE; 9228023.  
RA MALKIN D., JOLLY K.W., BARBIER N., LOOK A.T., FRIEND S.H.,  
RA GERHARDT M.C., ANDERSEN T.I., BORRESEN A.-L., LI F.P., GARBER J.,  
RA STRONG L.C.;  
RT "Germline mutations of the p53 tumor-suppressor gene in children and  
young adults with second malignant neoplasms.";  
RL New Engl. J. Med. 326:1309-1315(1992).  
[25]  
RP VARIANTS BREAST TUMORS GLN-132; SER-249; LYS-280 AND LYS-285.  
RX MEDLINE; 90295284.  
RA BARTEK J., IGGO R., GANNON J., LANE D.P.;  
RT "Genetic and immunochemical analysis of mutant p53 in human breast  
cancer cell lines.";  
RL Oncogene 5:893-899(1990).  
[26]  
RP VARIANTS COLON TUMORS PHE-241 AND HIS-273.  
RX MEDLINE; 91017544.  
RA RODRIGUES N.R., ROWAN A., SMITH M.E.F., KERR I.B., BODMER W.F.,  
RA GANNON J.V., LANE D.P.;  
...: remainder of annotations omitted.

Query Match 100.0%; Score 63; DB 1; Length 393;  
Best Local Similarity 100.0%; Pred. No. 3 34e-04;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 154 GTRVRAMAI 162

Qy 1 GTRVRAMAI 9

RESULT 9

ID P53\_CRIGR STANDARD; PRT; 393 AA.  
AC 009185; Q64397; P97258; P97788;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE CELLULAR TUMOR ANTIGEN P53.  
GN TP53 OR P53.  
OS Cricetus griseus (Chinese hamster).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae; Cricetus.  
[1]  
RP SEQUENCE FROM N.A.  
RA CHAUNG W., MI L.J., BOORSTEIN R.J.;  
RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.  
[2]  
RP SEQUENCE FROM N.A.  
RC TISSUE-LIVER;  
RX MEDLINE; 97183659.  
RA LEE H., LARNER J.M., HAMLIN J.L.;  
RT "Cloning and characterization of Chinese hamster p53 cDNA.";  
RL Gene 184:177-183(1997).  
[3]  
RP SEQUENCE FROM N.A.  
RC TISSUE-EMBRYONIC FIBROBLAST;  
RA SHIMIZU T., NIKAI O., SUZUKI F.;  
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES  
GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL  
CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN  
TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION, IT IS A  
TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION  
BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF  
THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.  
APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF  
BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2  
EXPRESSION.  
CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
CC -1- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY  
OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED  
IN MANY TYPES OF CANCER.  
CC -1- SIMILARITY: BELONGS TO THE P53 FAMILY.  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC EMBL; Y08900; CAA70108.1; -  
CC EMBL; Y08901; CAA70109.1; -  
CC EMBL; U50395; AAC53040.1; -  
CC EMBL; D86070; BAA13004.1; -  
CC HSP; P04637; IYCQ.  
CC PROSITE; PS00348; P53; 1.  
CC PFAM; PF00870; P53; 1.  
CC Anti-oncogene; DNA-binding; Transcription regulation; Activator;  
KW Nuclear protein; Phosphorylation; Apoptosis.  
FT DOMAIN 1 74 ASP/GLU-RICH (ACIDIC).  
FT DOMAIN 75 150 HYDROPHOBIC.  
FT DOMAIN 316 390 HIGHLY BASIC AND MAY BE INVOLVED IN  
INTERACTION WITH DNA.  
FT DOMAIN 311 323 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
FT MOD\_RES 392 392 PHOSPHORYLATION (BY SIMILARITY).  
FT VARIANT 133 133 L -> Q (IN CELL LINE V79-4).  
FT VARIANT 135 135 C -> W (IN CELL LINE V79-4).  
FT CONFLICT 103 Y -> F (IN REF. 2).  
SQ SEQUENCE 393 AA; 43378 MW; 402EB149 CRC32;  
Query Match 100.0%; Score 63; DB 1; Length 393;  
Best Local Similarity 100.0%; Pred. No. 3 34e-04;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Db 154 GTRVRAMAI 162
| | | | | | | |
Qy 1 GTRVRAMAI 9

RESULT 10
ID P53_MESAU STANDARD; PRT; 396 AA.
AC Q00366; P97276;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE CELLULAR TUMOR ANTIGEN P53.
GN TP53.
OS Mesocricetus auratus (Golden hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae; Mesocricetus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SYRIAN; TISSUE-KIDNEY;
RX MEDLINE; 92210007.
RA LEGROS Y., MCINTYRE P., SOUSSI T.;
RT "The CDNA cloning and immunological characterization of hamster p53.";
RL Gene 112:247-250(1992).
RN [2]
RP SEQUENCE FROM N.A.
RA HOU E.W., WISEMAN R.;
RL Submitted (APR-1994) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES
CC GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL
CC CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN
CC TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION, IT IS A
CC TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION
CC BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF
CC THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.
CC APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF
CC BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2
CC EXPRESSION.
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY
CC OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED
CC IN MANY TYPES OF CANCER.
CC -!- SIMILARITY: BELONGS TO THE P53 FAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announcement/
or send an email to license@isb-sib.ch).
-----
EMBL; M75144; AAA37085.1; -
EMBL; U07182; AAB41344.1; -
PIR; JH0633; JH0633.
HSSP; P04637; IYCQ.
PROSITE; PS00348; P53; 1.
PFAM; PF00870; P53; 1.
KW Anti-oncogene; DNA-binding; Transcription regulation; Activator;
Nuclear protein; Phosphorylation; Apoptosis.
FT DOMAIN 1 77 ASP/GLU-RICH (ACIDIC).
FT DOMAIN 78 153 HYDROPHOBIC.
FT FT 153
FT DOMAIN 319 393 INTERACTION WITH DNA
FT FT 314 326 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT MOD_RES 395 395 PHOSPHORYLATION (BY SIMILARITY).
FT CONFLICT 188 188 G -> S (IN REF. 2).
SQ SEQUENCE 396 AA; 43631 MW; C2668ADE CRC32;

Query Match 100.0%; Score 63; DB 1; Length 396;
Best Local Similarity 100.0%; Pred. No. 3.34e-04;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Db 157 GTRVRAMAI 165
| | | | | | | |
Qy 1 GTRVRAMAI 9

RESULT 11
ID P53_MOUSE STANDARD; PRT; 390 AA.
AC P02340;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE CELLULAR TUMOR ANTIGEN P53.
GN TP53 OR TRP53 OR P53.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 85027173.
RA BIENZ B., ZAKUT-HOURI R., GIVOL D., OREN M.;
RT "Analysis of the gene coding for the murine cellular tumour antigen
p53.";
RL EMBO J. 3:2179-2183(1984).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 84068204.
RA ZAKUT-HOURI R., OREN M., BIENZ B., LAVIE V., HAZUM S., GIVOL D.;
RT "A single gene and a pseudogene for the cellular tumour antigen p53.";
RL Nature 306:594-597(1983).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE; 84272240.
RA JENKINS J.R., RUDGE K., REDMOND S., WADE-EVANS A.;
RT "Cloning and expression analysis of full length mouse CDNA sequences
encoding the transformation associated protein p53.";
RL Nucleic Acids Res. 12:5609-5626(1984).
RN [4]
RP SEQUENCE FROM N.A. (CLONES PCD53; P53-M11 AND P53-M8).
RX MEDLINE; 87064640.
RA ARAI N., NOMURA D., YOKOTA K., WOLF D., BRILL E., SHOHAT O.,
RA ROTTER V.;
RT "Immunologically distinct p53 molecules generated by alternative
splicing.";
RL Mol. Cell. Biol. 6:3232-3239(1986).
RN [5]
RP SEQUENCE OF 222-258 FROM N.A.
RX MEDLINE; 92115342.
RA BURNS P.A., KEMP C.J., GANNON J.V., LANE D.P., BREMMER R.,
RA BALMAIN A.;
RT "Loss of heterozygosity and mutational alterations of the p53 gene in
skin tumours of interspecific hybrid mice.";
RL Oncogene 6:2363-2369(1991).
RN [6]
RP PHOSPHORYLATION SITES.
RX MEDLINE; 86149247.
RA SAWAD A., ANDERSON C.W., CARROLL R.B.;
RT "Mapping of phosphomonoester and apparent phosphodiester bonds of the
oncogene product p53 from simian virus 40-transformed 3T3 cells.";
RL Proc. Natl. Acad. Sci. U.S.A. 83:897-901(1986).
RN [7]
RP PHOSPHORYLATION SITES.
RX MEDLINE; 91006019.
RA MEEK D.W., SIMON S., KIKKAWA U., ECKHART W.;
RT "The p53 tumour suppressor protein is phosphorylated at serine 389 by
casein kinase II."
RL EMBO J. 9:3253-3260(1990).
CC -!- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES
CC GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL
CC CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN
CC TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION, IT IS A
CC TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION
CC BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF
CC THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.

```



```

CC Macaca.
RN [1]
RA SEQUENCE FROM N.A.
RP KHAN M.A., HANSEN C., WELSH J.A., BENNETT W.P.;
RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES
CC GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL
CC CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN
CC TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION, IT IS A
CC TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION
CC BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF
CC THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.
CC APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF
CC BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2
CC EXPRESSION.
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY
CC OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED
CC IN MANY TYPES OF CANCER.
CC -!- SIMILARITY: BELONGS TO THE P53 FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U48956; AAB91534.1; -
CC HSSP; P04637; ISAH.
CC PROSITE; PS00348; P53; 1.
CC PFAM; PF00870; P53; 1.
CC Anti-oncogene; DNA-binding; Transcription regulation; Activator;
CC Nuclear protein; Phosphorylation; Apoptosis.
CC DOMAIN 1 80 ASP/GLU-RICH (ACIDIC).
CC DOMAIN 81 150 HYDROPHOBIC.
CC DOMAIN 319 393 HIGHLY BASIC AND MAY BE INVOLVED IN
CC INTERACTION WITH DNA.
CC DOMAIN 311 323 NUCLEAR LOCALIZATION SIGNAL.
CC MOD_RES 392 392 PHOSPHORYLATION (BY SIMILARITY).
CC SEQUENCE 393 AA; 43655 MW; 11A9B7F8 CRC32;
CC -----
CC Query Match 95.2%; Score 60; DB 1; Length 393;
CC Best Local Similarity 88.9%; Pred. No. 2.27e-03;
CC Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
CC -----
Db 154 GSRVRAMAI 162
QY 1 GTRVRAMAI 9
|:|||||||
RESULT 14
ID P53_MACFA STANDARD; PRT; 393 AA.
AC P56423;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE CELLULAR TUMOR ANTIGEN P53.
GN TP53 OR P53.
OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Cercopithecoidea; Cercopithecinae;
OC Macaca.
CC [1]
CC SEQUENCE FROM N.A.
CC KHAN M.A., HANSEN C., WELSH J.A., BENNETT W.P.;
CC Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES
CC GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL
CC CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN
CC TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION, IT IS A
CC TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION
CC BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF
CC THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.
CC APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF
CC BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2
CC EXPRESSION.
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY
CC OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED
CC IN MANY TYPES OF CANCER.
CC -!- SIMILARITY: BELONGS TO THE P53 FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U48956; AAB91534.1; -
CC HSSP; P04637; ISAH.
CC PROSITE; PS00348; P53; 1.
CC PFAM; PF00870; P53; 1.
CC Anti-oncogene; DNA-binding; Transcription regulation; Activator;
CC Nuclear protein; Phosphorylation; Apoptosis.
CC DOMAIN 1 80 ASP/GLU-RICH (ACIDIC).
CC DOMAIN 81 150 HYDROPHOBIC.
CC DOMAIN 319 393 HIGHLY BASIC AND MAY BE INVOLVED IN
CC INTERACTION WITH DNA.
CC DOMAIN 311 323 NUCLEAR LOCALIZATION SIGNAL.
CC MOD_RES 392 392 PHOSPHORYLATION (BY SIMILARITY).
CC SEQUENCE 393 AA; 43655 MW; 11A9B7F8 CRC32;
CC -----
CC Query Match 95.2%; Score 60; DB 1; Length 393;
CC Best Local Similarity 88.9%; Pred. No. 2.27e-03;
CC Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
CC -----
Db 154 GSRVRAMAI 162
QY 1 GTRVRAMAI 9
|:|||||||
RESULT 15
ID RPOE_SULAC STANDARD; PRT; 248 AA.
AC P39466;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-FEB-1995 (Rel. 31, Last annotation update)
DE DNA-DIRECTED RNA POLYMERASE SUBUNIT E (EC 2.7.7.6).
GN RPOE.
OS Sulfolobus acidocaldarius.
OC Archaea; Crenarchaeota; Sulfolobales; Sulfolobus.
CC [1]
CC SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
CC STRAIN-DSM 639;
CC MEDLINE; 94173739.
CC LANGER D., LOTTSCHEICH F., ZILLIG W.;
CC "A subunit of an archaeal DNA-dependent RNA polymerase contains the
CC S1 motif".
CC Nucleic Acids Res. 22:694-694(1994).
CC -!- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
CC SUBSTRATES.
CC -!- CATALYTIC ACTIVITY: N NUCLEOSIDE TRIPHOSPHATE -> N PYROPHOSPHATE +
CC RNA(N).
CC -!- SUBUNIT: THE S.ACIDOCALDIARIUS RNAP IS COMPOSED OF 13 SUBUNITS.
CC -!- SIMILARITY: CONTAINS A COPY OF THE 'S1 MOTIF'.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```

CC entitles requires a license agreement (See <http://www.isb-sib.ch/announce/>;  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; X75411; CAA53164.1; -.  
DR PIR; S38658; S38658.  
DR PIR; S42389; S42389.  
DR PFAM; PF00575; S1; 1.  
KW Transferase; Transcription; DNA-directed RNA polymerase; Zinc-finger.  
FT ZN\_FING 196 213 C4-TYPE.  
SQ SEQUENCE 248 AA; 27632 MW; AE1B2336 CRC32;

Query Match 82.5%; Score 52; DB 1; Length 248;  
Best Local Similarity 77.8%; Pred. No. 2.86e-01;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 139 GDRVAMII 147

Qy 1 GTRVAMAI 9

Search completed: Sat Apr 15 00:37:35 2000  
Job time : 40 secs.

\*\*\*\*\*  
[W][O][R][L][D]  
\*\*\*\*\*  
(TW)

Release 3.1a John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run On: Sat Apr 15 00:37:52 2000; MasPar time 7.07 Seconds  
88.232 Million cell updates/sec  
Tabular output not generated.

Title: >US-08-452-843-16  
Description: (1-9) from US08452843.pep  
Perfect Score: 63  
Sequence: 1 GTRVRAMAI 9

Scoring table: PAM 150  
Gap 15

Searched: 225878 seqs, 69334122 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: sptrembl12  
1:sp-archaea 2:sp-bacteria 3:sp-fungi 4:sp-human  
5:sp-invertebrate 6:sp-mammal 7:sp-mhc 8:sp-organelle  
9:sp-phage 10:sp-plant 11:sp-rodent 12:sp-unclassified  
13:sp-vertebrate 14:sp-virus

Statistics: Mean 22.556; Variance 23.885; scale 0.944

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %	Match	Length	ID	Description	Pred. No.
1	63	100.0	95	11	Q92010	TUMOR-SUPPRESSOR P53 (	8.58e-04
2	63	100.0	95	11	Q92011	TUMOR-SUPPRESSOR P53 (	8.58e-04
3	63	100.0	95	11	Q92012	TUMOR-SUPPRESSOR P53 (	8.58e-04
4	63	100.0	136	11	Q64396	CELLULAR TUMOR ANTIGEN	8.58e-04
5	63	100.0	136	11	Q64396	CELLULAR TUMOR ANTIGEN	8.58e-04
6	63	100.0	196	6	Q23484	CELLULAR TUMOR ANTIGEN	8.58e-04
7	63	100.0	205	11	Q35873	CELLULAR TUMOR ANTIGEN	8.58e-04
8	63	100.0	391	6	Q36006	CELLULAR TUMOR ANTIGEN	8.58e-04
9	63	100.0	393	4	Q16535	P53 TRANSFORMATION SUP	8.58e-04
10	63	100.0	393	4	Q15087	P53 TRANSFORMATION SUP	8.58e-04
11	63	100.0	393	4	Q15088	P53 TRANSFORMATION SUP	8.58e-04
12	63	100.0	393	4	Q15089	CELLULAR TUMOR ANTIGEN	8.58e-04
13	63	100.0	393	4	Q16808	CELLULAR TUMOR ANTIGEN	8.58e-04
14	63	100.0	393	4	Q15086	P53 TRANSFORMATION SUP	8.58e-04
15	63	100.0	393	4	Q15088	P53 TRANSFORMATION SUP	8.58e-04
16	63	100.0	393	4	Q15088	CELLULAR TUMOR ANTIGEN	8.58e-04
17	63	100.0	393	4	Q15810	CELLULAR TUMOR ANTIGEN	8.58e-04
18	63	100.0	393	4	Q15807	CELLULAR TUMOR ANTIGEN	8.58e-04
19	63	100.0	393	4	Q16811	CELLULAR TUMOR ANTIGEN	8.58e-04
20	60	95.2	90	11	P70656	P53 (FRAGMENT).	5.55e-03

21	60	95.2	135	11	Q64451	CELLULAR TUMOR ANTIGEN	5.55e-03
22	60	95.2	238	14	P89004	P53 (FRAGMENT).	5.55e-03
23	60	95.2	286	14	P90332	P53 (FRAGMENT).	5.55e-03
24	60	95.2	286	14	P89003	P53 (FRAGMENT).	5.55e-03
25	60	95.2	378	14	P89002	P53 (FRAGMENT).	5.55e-03
26	60	95.2	390	11	O70366	CELLULAR TUMOR ANTIGEN	5.55e-03
27	56	88.9	391	11	Q9WUR6	CELLULAR TUMOR ANTIGEN	6.20e-02
28	54	85.7	564	1	O26392	SENSORY TRANSDUCTION H	2.00e-01
29	53	84.1	114	1	Q9YBL8	114AA LONG HYPOTHETICA	3.56e-01
30	53	84.1	567	1	O26456	SENSORY TRANSDUCTION H	3.56e-01
31	50	79.4	386	9	Q38068	TAIL SHEATH PROTEIN (G	1.93e-00
32	49	77.8	495	1	O26559	SENSORY TRANSDUCTION H	3.33e-00
33	48	76.2	209	2	P71647	HYPOTHETICAL 22.5 KD P	5.72e-00
34	48	76.2	383	2	O07407	ALCOHOL DEHYDROGENASE	5.72e-00
35	48	76.2	397	2	O66036	SULFATE ADENYLYLTRANSF	5.72e-00
36	47	74.6	172	2	P94384	YCGI PROTEIN.	9.74e+00
37	47	74.6	244	2	O31473	YCGI PROTEIN.	9.74e+00
38	47	74.6	543	4	O60407	PAC CLONE DJ1168D11 FR	9.74e+00
39	47	74.6	602	3	P78739	EXCHITININASE.	9.74e+00
40	47	74.6	1209	5	P91581	COS41.3.	9.74e+00
41	46	73.0	167	2	O68719	ISI222 HYPOTHETICAL PR	1.65e+01
42	46	73.0	192	10	P93169	EARLY LIGHT-INDUCED PR	1.65e+01
43	46	73.0	413	2	O54482	TYPE IV PILUS ASSEMBLY	1.65e+01
44	46	73.0	536	2	Q923D5	POLY-DETA-HYDROXYBUTYR	1.85e+01
45	46	73.0	958	10	Q92QA2	HYPOTHETICAL 106.0 KD	1.65e+01

ALIGNMENTS

RESULT 1	PRELIMINARY;	PRT;	95 AA.
ID Q92010			
AC Q92010;			
DT 01-MAY-1999 (TREMBlrel. 10, Created)			
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)			
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)			
DE TUMOR-SUPPRESSOR P53 (FRAGMENT).			
GN P53.			
OS Microtus rossiaemeridionalis.			
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;			
OC Eutheria; Rodentia; Sciurognathi; Muridae; Arvicolinae; Microtus.			
RN [1]			
RC SEQUENCE FROM N.A.			
RP STRAIN-VARIOUS STRAINS;			
RA DEWOODY J.A.;			
RT "Nucleotide variation in the p53 tumor-suppressor gene of voles from Chernobyl, Ukraine."			
RL Mutat. Res. 0:0-0(1998).			
DR EMBL; AF014036; AAC78751.1; -			
DR EMBL; AF014017; AAC78732.1; -			
DR EMBL; AF014022; AAC78737.1; -			
DR EMBL; AF014024; AAC78739.1; -			
DR EMBL; AF014027; AAC78742.1; -			
DR EMBL; AF014028; AAC78743.1; -			
DR EMBL; AF014030; AAC78745.1; -			
DR EMBL; AF014033; AAC78748.1; -			
DR EMBL; AF014034; AAC78749.1; -			
DR HSSP; P04637; 1TSR.			
FT NON_TER 1			
FT NON_TER 95			
SQ SEQUENCE 95 AA; 10864 MW; 0BEE8D3A CRC32;			
Query Match 100.0%; Score 63; DB 11; Length 95;			
Best Local Similarity 100.0%; Pred. No. 8.58e-04;			
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
Db 1 GTRVRAMAI 9			
Qy 1 GTRVRAMAI 9			
RESULT 2	PRELIMINARY;	PRT;	95 AA.
ID Q92011			

```

AC Q92011;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
DE TUMOR-SUPPRESSOR P53 (FRAGMENT).
GN P53.
OS Microtus oeconomus.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Arvicolinae; Microtus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-TK44915, TK44840, TK44844;
RA DEWOODY J.A.;
RT "Nucleotide variation in the p53 tumor-suppressor gene of voles from
  Chernobyl, Ukraine.";
RL Mutat. Res. 0:0-0(1998).
DR EMBL; AF014045; AAC78760.1; -
DR EMBL; AF014043; AAC78758.1; -
DR EMBL; AF014044; AAC78759.1; -
DR HSSP; P04637; ITSR.
FT NON_TER 1
FT NON_TER 95
SQ SEQUENCE 95 AA; 10864 MW; 0BEE8D3A CRC32;

Query Match 100.0%; Score 63; DB 11; Length 95;
Best Local Similarity 100.0%; Pred. No. 8.58e-04;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 GTRVRAMAI 9
QY 1 GTRVRAMAI 9
|||||||

RESULT 3
ID Q922V1 PRELIMINARY; PRT; 95 AA.
AC Q922V1;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
DE TUMOR-SUPPRESSOR P53 (FRAGMENT).
GN P53.
OS Microtus agrestis (short-tailed field vole).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Arvicolinae; Microtus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-TK50054;
RA DEWOODY J.A.;
RT "Nucleotide variation in the p53 tumor-suppressor gene of voles from
  Chernobyl, Ukraine.";
RL Mutat. Res. 0:0-0(1998).
DR EMBL; AF014046; AAC78761.1; -
DR HSSP; P04637; ITSR.
FT NON_TER 1
FT NON_TER 95
SQ SEQUENCE 95 AA; 10864 MW; 0BEE8D3A CRC32;

Query Match 100.0%; Score 63; DB 11; Length 95;
Best Local Similarity 100.0%; Pred. No. 8.58e-04;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 GTRVRAMAI 9
QY 1 GTRVRAMAI 9
|||||||

RESULT 4
ID Q92012 PRELIMINARY; PRT; 95 AA.
AC Q92012;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
DE TUMOR-SUPPRESSOR P53 (FRAGMENT).

```

```

GN P53.
OS Microtus arvalis.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Arvicolinae; Microtus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-VARIOUS STRAINS;
RA DEWOODY J.A.;
RT "Nucleotide variation in the p53 tumor-suppressor gene of voles from
  Chernobyl, Ukraine.";
RL Mutat. Res. 0:0-0(1998).
DR EMBL; AF014047; AAC78762.1; -
DR EMBL; AF014018; AAC78733.1; -
DR EMBL; AF014019; AAC78734.1; -
DR EMBL; AF014020; AAC78735.1; -
DR EMBL; AF014021; AAC78736.1; -
DR EMBL; AF014023; AAC78738.1; -
DR EMBL; AF014025; AAC78740.1; -
DR EMBL; AF014026; AAC78741.1; -
DR EMBL; AF014029; AAC78744.1; -
DR EMBL; AF014031; AAC78746.1; -
DR EMBL; AF014032; AAC78747.1; -
DR EMBL; AF014037; AAC78752.1; -
DR EMBL; AF014038; AAC78753.1; -
DR EMBL; AF014039; AAC78754.1; -
DR EMBL; AF014040; AAC78755.1; -
DR EMBL; AF014041; AAC78756.1; -
DR EMBL; AF014042; AAC78757.1; -
DR HSSP; P04637; ITSR.
FT NON_TER 1
FT NON_TER 95
SQ SEQUENCE 95 AA; 10864 MW; 0BEE8D3A CRC32;

Query Match 100.0%; Score 63; DB 11; Length 95;
Best Local Similarity 100.0%; Pred. No. 8.58e-04;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 GTRVRAMAI 9
QY 1 GTRVRAMAI 9
|||||||

RESULT 5
ID Q64396 PRELIMINARY; PRT; 136 AA.
AC Q64396; P97940;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
DE CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).
GN P53.
OC Cricetulus griseus (Chinese hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae; Cricetulus.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-KIDNEY;
RA SIWASKI D., MAI S., SCHNEIDERMAN M.H., HUPPI K.;
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT
CC PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL
CC CYCLE REGULATION, IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY
CC REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED
CC FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF
CC CYCLIN-DEPENDENT KINASES (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
DR EMBL; U41451; AAB41266.1; -
DR HSSP; P04637; ITSR.
DR PROSITE; PS00348; P53; 1.
DR PFAM; PF00870; P53; 1.
KW Nuclear protein; Phosphorylation; Anti-oncogene; DNA-binding;
KW Transcription regulation; Activator.
FT NON_TER 1
FT NON_TER 136

```



SQ SEQUENCE 136 AA; 15411 MW; CTF916C9 CRC32;  
Query Match 100.0%; Score 63; DB 11; Length 136;  
Best Local Similarity 100.0%; Pred. No. 8.58e-04;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Db 23 GTRVRAMAI 31  
| | | | |  
QY 1 GTRVRAMAI 9  
RESULT 6  
ID Q60434 PRELIMINARY; PRT; 136 AA.  
AC Q60434; P97257;  
DT 01-NOV-1996 (TREMELrel. 01, Created)  
DT 01-JAN-1998 (TREMELrel. 05, Last sequence update)  
DT 01-NOV-1999 (TREMELrel. 12, Last annotation update)  
DE CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).  
OS Crictetus griseus (Chinese hamster).  
OC Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Mammalia;  
OC Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae; Crictetus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA SIWASKI D., MAI S., SCHNEIDERMAN M.H., HUPPI K.;  
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT  
CC PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL  
CC CYCLE REGULATION, IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY  
CC REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED  
CC FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF  
CC CYCLIN-DEPENDENT KINASES (BY SIMILARITY).  
CC -!- SUBCELLULAR LOCATION: NUCLEAR.  
DR EMBL; U41452; AAB41267.1; -.  
DR HSSP; P04637; 1TSR.  
DR PROSITE; PS00348; P53; 1.  
DR PFAM; PF00870; P53; 1.  
DR KW Nuclear oncogene; DNA-binding; Transcription regulation; Activator;  
FT NON\_TER 1  
FT NON\_TER 136  
SQ SEQUENCE 136 AA; 15438 MW; 10679AD4 CRC32;  
Query Match 100.0%; Score 63; DB 11; Length 136;  
Best Local Similarity 100.0%; Pred. No. 8.58e-04;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Db 23 GTRVRAMAI 31  
| | | | |  
QY 1 GTRVRAMAI 9  
RESULT 7  
ID Q29484 PRELIMINARY; PRT; 196 AA.  
AC Q29484;  
DT 01-NOV-1996 (TREMELrel. 01, Created)  
DT 01-NOV-1999 (TREMELrel. 01, Last sequence update)  
DT 01-NOV-1999 (TREMELrel. 12, Last annotation update)  
DE CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).  
GN P53.  
OS Equus caballus (Horse).  
OC Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Mammalia;  
OC Eutheria; Perissodactyla; Equidae; Equus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA BUCHER K., SZALAI G., MARTI E., PAULI U., LAZARY S.;  
RL Res. Vet. Sci. 0:0-0(0).  
CC -!- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT  
CC PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL  
CC CYCLE REGULATION, IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY  
CC REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED  
CC FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF  
CC CYCLIN-DEPENDENT KINASES (BY SIMILARITY).  
CC -!- SUBCELLULAR LOCATION: NUCLEAR.

DR EMBL; X91793; CAA62905.1; -.  
DR HSSP; P04637; 1SAH.  
DR PROSITE; PS00348; P53; 1.  
DR PFAM; PF00870; P53; 1.  
KW Anti-oncogene; DNA-binding; Transcription regulation; Activator;  
KW Nuclear protein; Phosphorylation.  
FT NON\_TER 1  
FT NON\_TER 196  
SQ SEQUENCE 196 AA; 22080 MW; F443239C CRC32;  
Query Match 100.0%; Score 63; DB 6; Length 196;  
Best Local Similarity 100.0%; Pred. No. 8.58e-04;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Db 20 GTRVRAMAI 28  
| | | | |  
QY 1 GTRVRAMAI 9  
RESULT 8  
ID O35873 PRELIMINARY; PRT; 205 AA.  
AC O35873;  
DT 01-JAN-1998 (TREMELrel. 05, Created)  
DT 01-JAN-1998 (TREMELrel. 05, Last sequence update)  
DT 01-NOV-1999 (TREMELrel. 12, Last annotation update)  
DE CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).  
GN P53.  
OS Crictetus griseus (Chinese hamster).  
OC Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Mammalia;  
OC Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae; Crictetus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA RAINALDI G., MARCHETTI S., CAPECCHI B., MENEVERI R., PIRAS A.,  
RA LEUZZI R.;  
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA VATERONI L., MUSIO A., MENEVERI R., RAINALDI G.;  
RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT  
CC PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL  
CC CYCLE REGULATION, IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY  
CC REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED  
CC FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF  
CC CYCLIN-DEPENDENT KINASES (BY SIMILARITY).  
CC -!- SUBCELLULAR LOCATION: NUCLEAR.  
DR EMBL; U74487; AAB82420.1; -.  
DR HSSP; P04637; 1SAH.  
DR PROSITE; PS00348; P53; 1.  
DR PFAM; PF00870; P53; 1.  
KW Anti-oncogene; DNA-binding; Transcription regulation; Activator;  
KW Nuclear protein; Phosphorylation.  
FT NON\_TER 1  
FT NON\_TER 205  
SQ SEQUENCE 205 AA; 23122 MW; 680DDDDC CRC32;  
Query Match 100.0%; Score 63; DB 11; Length 205;  
Best Local Similarity 100.0%; Pred. No. 8.58e-04;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Db 29 GTRVRAMAI 37  
| | | | |  
QY 1 GTRVRAMAI 9  
RESULT 9  
ID O36006 PRELIMINARY; PRT; 391 AA.  
AC O36006;  
DT 01-JAN-1998 (TREMELrel. 05, Created)  
DT 01-JAN-1998 (TREMELrel. 05, Last sequence update)  
DT 01-NOV-1999 (TREMELrel. 12, Last annotation update)  
DE CELLULAR TUMOR ANTIGEN P53.  
GN P53.

OS Marmota monax (Woodchuck).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Rodentia; Sciurognathi; Scuridae; Scurinae; Marmota.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 97376996.  
 RA FEITELSON M.A., RANGANATHAN P.N., CLAYTON M.M., ZHANG S.M.;  
 RT "Partial characterization of the woodchuck tumor suppressor, p53, and  
 RT its interaction with woodchuck hepatitis virus X antigen in  
 RT hepatocarcinogenesis.";  
 RL Oncogene 15:327-336(1997).  
 CC -1- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT  
 CC PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL  
 CC CYCLE REGULATION, IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY  
 CC REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED  
 CC FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF  
 CC CYCLIN-DEPENDENT KINASES (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 DR EMBL; AJ001022; CAA04478.1; -.  
 DR HSSP; P04637; ITSR.  
 DR PROSITE; PS00348; P53; 1.  
 DR PFAM; PF00870; P53; 1.  
 DR PRINTS; PR00386; P53SUPPRESSR.  
 KW Anti-oncogene; DNA-binding; Transcription regulation; Activator;  
 KW Nuclear protein; Phosphorylation.  
 SQ SEQUENCE 391 AA; 43468 MW; 95FAB8F2 CRC32;

Query Match 100.0%; Score 63; DB 6; Length 391;  
 Best Local Similarity 100.0%; Pred. No. 8.58e-04;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 152 GTRVRAMAI 160  
 |||||  
 QY 1 GTRVRAMAI 9

RESULT 10  
 ID Q16535 PRELIMINARY; PRT; 393 AA.  
 AC Q16535;  
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
 DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)  
 DE P53 TRANSFORMATION SUPPRESSOR (FRAGMENT).  
 GN P53.

OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 92007731.

RA FARRELL P.J., ALLAN G., SHANAHAN F., VOUSDEN K.H., CROOK T.;  
 RT "p53 is frequently mutated in Burkitt's lymphoma cell lines.";  
 RL EMO J. 10:2879-2887(1991).  
 DR EMBL; X60017; CAA42632.1; -.  
 DR EMBL; X60015; CAA42630.1; -.  
 DR HSSP; P04637; ISAH.  
 DR PFAM; PF00870; P53; 1.  
 DR VARIANT 248 248 Q -> R.  
 FT NON\_TER 393 393  
 SQ SEQUENCE 393 AA; 43684 MW; 239818A9 CRC32;

Query Match 100.0%; Score 63; DB 4; Length 393;  
 Best Local Similarity 100.0%; Pred. No. 8.58e-04;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 154 GTRVRAMAI 162  
 |||||  
 QY 1 GTRVRAMAI 9

RESULT 11  
 ID Q15087 PRELIMINARY; PRT; 393 AA.  
 AC Q15087;

DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
 DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)  
 DE P53 TRANSFORMATION SUPPRESSOR (FRAGMENT).  
 GN P53.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 92007731.  
 RA FARRELL P.J., ALLAN G., SHANAHAN F., VOUSDEN K.H., CROOK T.;  
 RT "p53 is frequently mutated in Burkitt's lymphoma cell lines.";  
 RL EMO J. 10:2879-2887(1991).  
 DR EMBL; X60014; CAA42629.1; -.  
 DR HSSP; P04637; ISAH.  
 DR PFAM; PF00870; P53; 1.  
 DR VARIANT 237 237 I -> M.  
 FT NON\_TER 393 393  
 SQ SEQUENCE 393 AA; 43694 MW; 9B81992 CRC32;

Query Match 100.0%; Score 63; DB 4; Length 393;  
 Best Local Similarity 100.0%; Pred. No. 8.58e-04;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 154 GTRVRAMAI 162  
 |||||  
 QY 1 GTRVRAMAI 9

RESULT 12  
 ID Q16809 PRELIMINARY; PRT; 393 AA.  
 AC Q16809;  
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
 DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)  
 DE CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).  
 GN P53.

OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 92007731.

RA FARRELL P.J., ALLAN G., SHANAHAN F., VOUSDEN K.H., CROOK T.;  
 RT "p53 is frequently mutated in Burkitt's lymphoma cell lines.";  
 RL EMO J. 10:2879-2887(1991).  
 CC -1- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT  
 CC PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL  
 CC CYCLE REGULATION, IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY  
 CC REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED  
 CC FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF  
 CC CYCLIN-DEPENDENT KINASES (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 DR EMBL; X60019; CAA42634.1; -.  
 DR HSSP; P04637; ISAH.  
 DR PROSITE; PS00348; P53; 1.  
 DR PFAM; PF00870; P53; 1.  
 DR Anti-oncogene; DNA-binding; Transcription regulation; Activator;  
 KW Nuclear protein; Phosphorylation.  
 FT VARIANT 213 213 Q -> R.  
 FT NON\_TER 393 393  
 SQ SEQUENCE 393 AA; 43684 MW; CB70BD7F CRC32;

Query Match 100.0%; Score 63; DB 4; Length 393;  
 Best Local Similarity 100.0%; Pred. No. 8.58e-04;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 154 GTRVRAMAI 162  
 |||||  
 QY 1 GTRVRAMAI 9

```
RESULT 13
ID Q16808 PRELIMINARY; PRT; 393 AA.
AC Q16808;
DT 01-NOV-1996 (TREMELREL. 01, Created)
DT 01-NOV-1996 (TREMELREL. 01, Last sequence update)
DT 01-NOV-1999 (TREMELREL. 12, Last annotation update)
DE CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).
GN P53.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 92007731.
RA FARRELL P.J., ALLAN G., SHANAHAN F., VOUSDEN K.H., CROOK T.;
RT "P53 is frequently mutated in Burkitt's lymphoma cell lines.";
RL EMBO J. 10:2879-2887(1991).
CC -1- FUNCTION: P53 SEEMS TO ACT AS A TUMOR SUPPRESSOR IN SOME, BUT
CC PROBABLY NOT ALL, TUMOR TYPES. P53 IS PROBABLY INVOLVED IN CELL
CC CYCLE REGULATION, IT IS A TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY
CC REGULATE CELLULAR DIVISION BY CONTROLLING A SET OF GENES REQUIRED
CC FOR THIS PROCESS. ONE OF THE GENES ACTIVATED IS AN INHIBITOR OF
CC CYCLIN-DEPENDENT KINASES (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
DR EMBL; X60018; CAA42633.1; -.
DR HSSP; P04637; 1SAH.
DR PROSITE; PS00348; P53; 1.
DR PFAM; PF00870; P53; 1.
KW Anti-oncogene; DNA-binding; Transcription regulation; Activator;
FT VARIANT 163 163 H -> Y.
FT NON_TER 393 393
SQ SEQUENCE 393 AA; 43627 MW; AFD8A9E3 CRC32;

Query Match 100.0%; Score 63; DB 4; Length 393;
Best Local Similarity 100.0%; Pred. No. 8.58e-04;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 154 GTRVRAMAI 162
|||||
QY 1 GTRVRAMAI 9

RESULT 14
ID Q15086 PRELIMINARY; PRT; 393 AA.
AC Q15086;
DT 01-NOV-1996 (TREMELREL. 01, Created)
DT 01-NOV-1996 (TREMELREL. 01, Last sequence update)
DT 01-NOV-1999 (TREMELREL. 12, Last annotation update)
DE P53 TRANSFORMATION SUPPRESSOR (FRAGMENT).
GN P53.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 92007731.
RA FARRELL P.J., ALLAN G., SHANAHAN F., VOUSDEN K.H., CROOK T.;
RT "P53 is frequently mutated in Burkitt's lymphoma cell lines.";
RL EMBO J. 10:2879-2887(1991).
DR EMBL; X60013; CAA42628.1; -.
DR HSSP; P04637; 1SAH.
DR PFAM; PF00870; P53; 1.
DR VARIANT 245 246 T -> M.
FT NON_TER 393 393
SQ SEQUENCE 393 AA; 43682 MW; 943B62A3 CRC32;

Query Match 100.0%; Score 63; DB 4; Length 393;
Best Local Similarity 100.0%; Pred. No. 8.58e-04;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 154 GTRVRAMAI 162
|||||
QY 1 GTRVRAMAI 9

RESULT 15
ID Q15088 PRELIMINARY; PRT; 393 AA.
AC Q15088;
DT 01-NOV-1996 (TREMELREL. 01, Created)
DT 01-NOV-1996 (TREMELREL. 01, Last sequence update)
DT 01-NOV-1999 (TREMELREL. 12, Last annotation update)
DE P53 TRANSFORMATION SUPPRESSOR (FRAGMENT).
GN P53.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 92007731.
RA FARRELL P.J., ALLAN G., SHANAHAN F., VOUSDEN K.H., CROOK T.;
RT "P53 is frequently mutated in Burkitt's lymphoma cell lines.";
RL EMBO J. 10:2879-2887(1991).
DR EMBL; X60016; CAA42631.1; -.
DR HSSP; P04637; 1SAH.
DR PFAM; PF00870; P53; 1.
DR VARIANT 238 238 Y -> C.
FT NON_TER 393 393
SQ SEQUENCE 393 AA; 43713 MW; A01E1523 CRC32;

Query Match 100.0%; Score 63; DB 4; Length 393;
Best Local Similarity 100.0%; Pred. No. 8.58e-04;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 154 GTRVRAMAI 162
|||||
QY 1 GTRVRAMAI 9

Search completed: Sat Apr 15 00:39:27 2000
Job time : 95 secs.
```

**THIS PAGE BLANK (USPTO)**

\*\*\*\*\*

\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

```

protein - protein database search, using Smith-Waterman algorithm
MParch_pp
Run on: Sat Apr 15 00:42:38 2000; MasPar time 3.12 Seconds
68.236 Million cell updates/sec
Tabular output not generated.

```

```

Title:
Description:
Perfect Score:
Sequence:

```

Scoring table: PAM 150  
Gap 15

Searched: 188963 seqs, 23686106 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: a-geneseq36  
1:geneseqp

Statistics: Mean 16.401; Variance 51.055; scale 0.321

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Query			ID	Description	Pred. No.
	Score	Match	Length			
1	64	100.0	113	1 R51877	Human p53 amino acids	3.07e+00
2	64	100.0	157	1 R51878	Human p53 amino acids	3.07e+00
3	64	100.0	335	1 W28498	Human p53 protein vari	3.07e+00
4	64	100.0	337	1 W13962	Chimeric p53 protein.	3.07e+00
5	64	100.0	353	1 W28494	Human p53 protein vari	3.07e+00
6	64	100.0	359	1 W13960	Chimeric p53 protein.	3.07e+00
7	64	100.0	361	1 W13961	Chimeric p53 protein.	3.07e+00
8	64	100.0	361	1 W13958	Chimeric p53 protein.	3.07e+00
9	64	100.0	363	1 W28479	Human p53 protein vari	3.07e+00
10	64	100.0	363	1 W28480	Human p53 protein vari	3.07e+00
11	64	100.0	363	1 W13975	Modified p53 variant p	3.07e+00
12	64	100.0	363	1 W13959	Chimeric p53 protein.	3.07e+00
13	64	100.0	374	1 W28482	Human p53 protein vari	3.07e+00
14	64	100.0	374	1 W28481	Human p53 protein vari	3.07e+00
15	64	100.0	381	1 W28489	Human p53 protein vari	3.07e+00
16	64	100.0	381	1 W28490	Human p53 protein vari	3.07e+00
17	64	100.0	390	1 W2623	Mouse p53 protein.	3.07e+00
18	64	100.0	393	1 Y03191	Amino acid sequence of	3.07e+00
19	64	100.0	393	1 W94270	Human p53 protein.	3.07e+00
20	64	100.0	393	1 W69218	Human p53 mutant 1.	3.07e+00
21	64	100.0	393	1 W69217	Human wild-type p53 pr	3.07e+00
22	64	100.0	393	1 W57244	Human p53 protein SEQ	3.07e+00
23	64	100.0	393	1 W57242	Human p53 protein SEQ	3.07e+00

24	64	100.0	393	1	W57243	Human p53 protein SQ	3.07e+00
25	64	100.0	393	1	W57245	Human p53 protein SQ	3.07e+00
26	64	100.0	393	1	W13949	T284R modified human p	3.07e+00
27	64	100.0	393	1	W13948	Human wild-type p53 tu	3.07e+00
28	64	100.0	393	1	W13953	T284K modified human p	3.07e+00
29	64	100.0	393	1	W13958	Modified p53 variant p	3.07e+00
30	64	100.0	393	1	W13970	Modified p53 variant p	3.07e+00
31	64	100.0	393	1	W25155	Human p53 variant foun	3.07e+00
32	64	100.0	393	1	W05345	Human p53 mutant N239S	3.07e+00
33	64	100.0	393	1	R191933	Wild type p53 protein.	3.07e+00
34	64	100.0	393	1	W13969	Modified p53 variant p	3.07e+00
35	64	100.0	393	1	W02637	Human p53 tumour suppr	3.07e+00
36	64	100.0	393	1	W13978	Human tumour-derived p	3.07e+00
37	64	100.0	393	1	W13952	Human tumour-derived p	3.07e+00
38	64	100.0	393	1	W13951	Human tumour-derived p	3.07e+00
39	64	100.0	393	1	W05349	Human p53 mutant R273C	3.07e+00
40	64	100.0	401	1	W28488	Human p53 protein vari	3.07e+00
41	64	100.0	402	1	W13955	Chimeric p53 protein.	3.07e+00
42	64	100.0	404	1	W13963	Chimeric p53 protein.	3.07e+00
43	64	100.0	406	1	W13966	Chimeric p53 protein.	3.07e+00
44	64	100.0	406	1	W13964	Chimeric p53 protein.	3.07e+00
45	64	100.0	411	1	W13967	Chimeric p53 protein.	3.07e+00

## ALIGNMENTS

RESULT	1	
ID	R51877 standard; Protein; 113 AA.	
AC	R51877;	
DE	18-NOV-1994 (first entry)	
DT	Human p53 amino acids 237-349.	
DE	Human nuclear phosphoprotein p53; tumour suppressor gene product;	
KW	anti-oncogene; cancer; tumour; antibody binding region; epitope.	
OS	Homo sapiens.	
FH	Key	Location/Qualifiers
FT	misc_difference 37	
FT	/note= "Arg corresponds to a CAT codon"	
PN	W09408241-A.	
PD	14-APR-1994.	
PF	30-SEP-1993; E02666.	
PR	30-SEP-1992; DE-232823.	
PA	(DEKR-) DEUT KREBSFORSCHUNGSZENTRUM STIFTUNG.	
PI	Klein R, Schranz P, Tesserer C, Voikmann M, Zentgraf H;	
DR	WPI; 94-135732/16.	
DR	N-PSDB; Q62362.	
PT	Non-radioactive detection of p53 specific antibodies - by capture	
PT	on immobilised p53 or its fragments, then reaction with labelled	
PT	second antibody, for diagnosis of tumours and suitable for	
PT	screening	
PS	Claim 10; Page 19; 35pp; German.	
CC	Antibodies specific for p53 are detected by binding to immobilised	
CC	fragments of the p53 gene product containing the antibody-binding	
CC	region. Preferred fragments contain amino acids 1-241, 40-349,	
CC	40-393, 66-241, 66-393, 237-349, 237-393 and esp. 9-33, 37-52 or	
CC	368-386. See R51872-R51881 for sequences of these fragments.	
SO	Sequence 113 AA;	

RESULT 2  
ID R51878 standard; Protein; 157 AA.  
AC R51878;  
DT 18-NOV-1994 (first entry)  
DE Human p53 amino acids 237-393.  
KW Human nuclear phosphoprotein p53;  
KW anti-oncogene; cancer; tumour;  
KW antibody binding region; epitope.

OS Homo sapiens. Location/Qualifiers  
FH Key misc\_difference 37  
FT /note= "Arg corresponds to a CAT codon"  
PN WO9408241-A.  
PD 14-APR-1994.  
PE 30-SEP-1993; DE-232823.  
PR PA (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM STIFTUNG.  
PI Klein R, Schranz P, Tesser C, Volkman M, Zentgraf H;  
DR WPI: 94-135732/16.  
DR N-PSDB: Q62363.  
PT Non-radioactive detection of p53 specific antibodies - by capture  
PT on immobilised p53 or its fragments, then reaction with labelled  
PT second antibody, for diagnosis of tumours and suitable for  
PT screening  
PS Claim 10; Page 19; 35pp; German.  
CC Antibodies specific for p53 are detected by binding to immobilised  
CC fragments of the p53 gene product containing the antibody-binding  
CC region. Preferred fragments contain amino acids 1-241, 40-349,  
CC 40-393, 66-241, 66-393, 237-349, 237-393 and esp. 9-33, 37-52 or  
CC 368-386. See R51872-R51881 for sequences of these fragments.  
SQ Sequence 157 AA;  
Query Match 100.0%; Score 64; DB 1; Length 157;  
Best Local Similarity 100.0%; Pred. No. 3.07e+00;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Db 13 RPILTIITL 21  
QY 1 RPILTIITL 9  
RESULT 3  
ID W28498 standard; Protein; 335 AA.  
AC W28498;  
DT 25-NOV-1997 (first entry)  
DE Human p53 protein variant 360h-325H.  
KW Leucine zipper domain; LZD; oligomerisation domain; mutant; mutein;  
KW substitution; replacement; transactivation; hinge region;  
KW anti-oncogene; hyperproliferation; cancer; restenosis;  
KW tumour suppression; apoptosis.  
OS Homo sapiens.  
OS Synthetic.  
FH Key Location/Qualifiers  
FT region 39..53  
FT /label= hinge  
FT misc\_difference 161  
FT /note= "Arg residue at position 182 of wild-type  
p53 has been mutated to His"  
PN WO9704092-A1.  
PD 06-FEB-1997.  
PE 17-JUL-1996; F01111.  
PR PA (RHON ) RHONE POULENC RORER SA.  
PI Bracco L, Conseiller E;  
DR WPI: 97-132633/12.  
PT New p53 variants e.g. with oligomerisation domain replaced by  
PT leucine zipper - useful for treating hyper-proliferative disorders,  
PT esp. cancer and restenosis  
PS Claim 39; Page -; 133pp; French.  
CC Claimed variants of protein p53 have at least part of the  
CC oligomerisation domain deleted and replaced by a leucine zipper  
CC domain. The mutants preferably also have at least part of the p53  
CC transactivation domain (amino acids 1-74) deleted and replaced by  
CC the domain 325-360 of p53. The present sequence is that of a  
CC specifically claimed p53 variant designated 360h-325H and comprising  
CC the 325-360 domain, separated from amino acids 75-325 of human  
CC wild-type p53 (but with Arg182 replaced by His) by a synthetic hinge  
CC sequence (Gly4Ser)3, and with a leucine zipper domain at the C-terminal.  
CC The p53 variants are more active and more stable tumour suppressors  
CC and apoptosis-inducing agents than wild-type p53 and are active where  
CC the wild-type protein is not, i.e. they are not inactivated by dominant

CC negative or oncogenic mutants, nor by other cellular proteins (because  
CC the leucine zipper domain prevents formation of inactive mixed  
CC oligomers).  
CC (Note: this sequence does not appear in the specification and has  
CC been produced by modifying the given sequence of variant 360h-325).  
SQ Sequence 335 AA;  
Query Match 100.0%; Score 64; DB 1; Length 335;  
Best Local Similarity 100.0%; Pred. No. 3.07e+00;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Db 229 RPILTIITL 237  
QY 1 RPILTIITL 9  
RESULT 4  
ID W13962 standard; Protein; 337 AA.  
AC W13962;  
DT 25-JUN-1997 (first entry)  
DE Chimeric p53 protein.  
KW p53; tumour suppressor; cancer; therapy; cell proliferation;  
KW apoptosis; protein engineering; GCN4; DNA binding.  
OS Chimeric Homo sapiens;  
OS Chimeric synthetic.  
FH Key Location/Qualifiers  
FT region 1..300  
FT /label= p53wt  
FT /note= "amino acids 1-300 of wild-type p53"  
FT region 301..305  
FT /label= Linker  
FT region 306..337  
FT /label= GCN4  
FT /note= "amino acids 250-281 of GCN4 LZ variant"  
PN WO9710843-A1.  
PD 27-MAR-1997.  
PE 20-SEP-1996; U15188.  
PR 22-SEP-1995; US-004802.  
PR 21-AUG-1996; US-697221.  
PA (WIST-) WISTAR INST ANATOMY & BIOLOGY.  
PI Halazonetis TD;  
DR WPI: 97-202618/18.  
PT R284K modified p53 protein having DNA binding ability - useful in  
PT treatment of cancer  
PS Disclosure: Refer to Page 8; 82pp; English.  
CC Chimeric p53 constructs (W13956-67) comprise N-terminal portions  
CC of human wild-type p53 tumour suppressor (see also W13948) linked  
CC to a C-terminal portion of the LZ variant (see also W13955) of  
CC GCN4 and, in some cases, the C-terminal portion of wild-type  
CC p53. The chimeric proteins have DNA binding activity and can  
CC replace lost or insufficient p53 function, providing the means for  
CC pharmacological rescue of p53 function in cancer patients. Nucleic  
CC acids coding for modified p53 constructs can be used for cancer  
CC gene therapy.  
SQ Sequence 337 AA;  
Query Match 100.0%; Score 64; DB 1; Length 337;  
Best Local Similarity 100.0%; Pred. No. 3.07e+00;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Db 249 RPILTIITL 257  
QY 1 RPILTIITL 9  
RESULT 5  
ID W28494 standard; Protein; 353 AA.  
AC W28494;  
DT 25-NOV-1997 (first entry)  
DE Human p53 protein variant 393-325H.  
KW Leucine zipper domain; LZD; oligomerisation domain; mutant; mutein;  
KW substitution; replacement; transactivation; viral protein VP16; HSV;  
KW anti-oncogene; hyperproliferation; cancer; restenosis.

KW tumour suppression; apoptosis.  
OS Homo sapiens.  
OS Synthetic.  
FH Key Location/Qualifiers  
FT misc\_difference 179  
FT /note= "Arg residue at position 182 of wild-type  
FT p53 has been mutated to His"  
PN WO9704092-A1.  
PD 06-FEB-1997.  
PF 17-JUL-1996; F01111.  
PR 19-JUL-1995; FR-008729.  
PA (RHON ) RHONE-POULENC RORER SA.  
PI Bracco L, Conseiller E;  
DR WPI; 97-132633/12.  
PT New p53 variants e.g. with oligomerisation domain replaced by  
PT leucine zipper - useful for treating hyper-proliferative disorders,  
PT esp. cancer and restenosis  
PS Claim 37; Page -; 133pp; French.  
CC Claimed variants of protein p53 have at least part of the  
CC oligomerisation domain deleted and replaced by a leucine zipper  
CC domain. The mutants preferably also have at least part of the p53  
CC transactivation domain (amino acids 1-74) deleted and replaced by  
CC the domain 325-393 of p53. The present sequence is that of a  
CC specifically claimed p53 variant designated 393-325H and comprising  
CC the 325-393 domain, amino acids 75-325 of human wild-type p53 (but with  
CC Arg182 replaced by His) and a leucine zipper domain at the C-terminal.  
CC The p53 variants are more active and more stable tumour suppressors  
CC and apoptosis-inducing agents than wild-type p53 and are active where  
CC the wild-type protein is not, i.e. they are not inactivated by dominant  
CC negative or oncogenic mutants, nor by other cellular proteins (because  
CC the leucine zipper domain prevents formation of inactive mixed  
CC oligomers).  
CC (Note: this sequence does not appear in the specification and has  
CC been produced by modifying the given sequence of variant 393-325).  
SQ Sequence 353 AA;  
Query Match 100.0%; Score 64; DB 1; Length 353;  
Best Local Similarity 100.0%; Pred. No. 3.07e+00;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Db 247 RPILTIITL 255  
|||||  
QY 1 RPILTIITL 9  
RESULT 6  
ID W13960 standard; Protein; 359 AA.  
AC W13960;  
DE 25-JUN-1997 (first entry)  
DE Chimeric p53 protein.  
KW p53; tumour suppressor; cancer; therapy; cell proliferation;  
KW apoptosis; protein engineering; GCN4; DNA binding.  
OS Chimeric Homo sapiens;  
OS Chimeric synthetic.  
FH Key Location/Qualifiers  
FT region 1..323  
FT /label= p53wt  
FT /note= "amino acids 1-323 of wild-type p53"  
FT region 324..326  
FT /label= Linker  
FT region 327..359  
FT /label= GCN4  
FT /note= "amino acids 249-281 of GCN4 LZ variant"  
PN WO9710843-A1.  
PD 27-MAR-1997.  
PF 20-SEP-1996; U15188.  
PR 22-SEP-1995; US-004802.  
PR 21-AUG-1996; US-697221.  
PA (WIST-) WISTAR INST ANATOMY & BIOLOGY.  
PI Halazonetis TD;  
DR WPI; 97-202618/18.  
PT R284K modified p53 protein having DNA binding ability - useful in  
PT treatment of cancer  
PS Disclosure; Refer to Page 8; 82pp; English.  
CC Chimeric p53 constructs (W13956-67) comprise N-terminal portions  
CC of human wild-type p53 tumour suppressor (see also W13948) linked  
CC to a C-terminal portion of the LZ variant (see also W13955) of  
CC GCN4 and, in some cases, the C-terminal portion of wild-type  
CC p53. The chimeric proteins have DNA binding activity and can  
CC replace lost or insufficient p53 function, providing the means for  
CC pharmacological rescue of p53 function in cancer patients. Nucleic  
CC acids coding for modified p53 constructs can be used for cancer  
CC gene therapy.  
SQ Sequence 359 AA;  
Query Match 100.0%; Score 64; DB 1; Length 361;  
Best Local Similarity 100.0%; Pred. No. 3.07e+00;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Db 249 RPILTIITL 257  
|||||  
QY 1 RPILTIITL 9  
RESULT 7  
ID W13961 standard; Protein; 361 AA.  
AC W13961;  
DE 23-JUN-1997 (first entry)  
DE Chimeric p53 protein.  
KW p53; tumour suppressor; cancer; therapy; cell proliferation;  
KW apoptosis; protein engineering; GCN4; DNA binding.  
OS Chimeric Homo sapiens;  
OS Chimeric synthetic.  
FH Key Location/Qualifiers  
FT region 1..323  
FT /label= p53wt  
FT /note= "amino acids 1-323 of wild-type p53"  
FT region 324..329  
FT /label= Linker  
FT region 330..361  
FT /label= GCN4  
FT /note= "amino acids 250-281 of GCN4 LZ variant"  
PN WO9710843-A1.  
PD 27-MAR-1997.  
PF 20-SEP-1996; U15188.  
PR 22-SEP-1995; US-004802.  
PR 21-AUG-1996; US-697221.  
PA (WIST-) WISTAR INST ANATOMY & BIOLOGY.  
PI Halazonetis TD;  
DR WPI; 97-202618/18.  
PT R284K modified p53 protein having DNA binding ability - useful in  
PT treatment of cancer  
PS Disclosure; Refer to Page 8; 82pp; English.  
CC Chimeric p53 constructs (W13956-67) comprise N-terminal portions  
CC of human wild-type p53 tumour suppressor (see also W13948) linked  
CC to a C-terminal portion of the LZ variant (see also W13955) of  
CC GCN4 and, in some cases, the C-terminal portion of wild-type  
CC p53. The chimeric proteins have DNA binding activity and can  
CC replace lost or insufficient p53 function, providing the means for  
CC pharmacological rescue of p53 function in cancer patients. Nucleic  
CC acids coding for modified p53 constructs can be used for cancer  
CC gene therapy.  
SQ Sequence 361 AA;  
Query Match 100.0%; Score 64; DB 1; Length 361;  
Best Local Similarity 100.0%; Pred. No. 3.07e+00;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Db 249 RPILTIITL 257  
|||||  
QY 1 RPILTIITL 9  
RESULT 8  
ID W13958 standard; Protein; 361 AA.  
AC W13958;

PS Disclosure; Refer to Page 8; 82pp; English.  
CC Chimeric p53 constructs (W13956-67) comprise N-terminal portions  
CC of human wild-type p53 tumour suppressor (see also W13948) linked  
CC to a C-terminal portion of the LZ variant (see also W13955) of  
CC GCN4 and, in some cases, the C-terminal portion of wild-type  
CC p53. The chimeric proteins have DNA binding activity and can  
CC replace lost or insufficient p53 function, providing the means for  
CC pharmacological rescue of p53 function in cancer patients. Nucleic  
CC acids coding for modified p53 constructs can be used for cancer  
CC gene therapy.  
SQ Sequence 359 AA;  
Query Match 100.0%; Score 64; DB 1; Length 359;  
Best Local Similarity 100.0%; Pred. No. 3.07e+00;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Db 249 RPILTIITL 257  
|||||  
QY 1 RPILTIITL 9  
RESULT 7  
ID W13961 standard; Protein; 361 AA.  
AC W13961;  
DE 23-JUN-1997 (first entry)  
DE Chimeric p53 protein.  
KW p53; tumour suppressor; cancer; therapy; cell proliferation;  
KW apoptosis; protein engineering; GCN4; DNA binding.  
OS Chimeric Homo sapiens;  
OS Chimeric synthetic.  
FH Key Location/Qualifiers  
FT region 1..323  
FT /label= p53wt  
FT /note= "amino acids 1-323 of wild-type p53"  
FT region 324..329  
FT /label= Linker  
FT region 330..361  
FT /label= GCN4  
FT /note= "amino acids 250-281 of GCN4 LZ variant"  
PN WO9710843-A1.  
PD 27-MAR-1997.  
PF 20-SEP-1996; U15188.  
PR 22-SEP-1995; US-004802.  
PR 21-AUG-1996; US-697221.  
PA (WIST-) WISTAR INST ANATOMY & BIOLOGY.  
PI Halazonetis TD;  
DR WPI; 97-202618/18.  
PT R284K modified p53 protein having DNA binding ability - useful in  
PT treatment of cancer  
PS Disclosure; Refer to Page 8; 82pp; English.  
CC Chimeric p53 constructs (W13956-67) comprise N-terminal portions  
CC of human wild-type p53 tumour suppressor (see also W13948) linked  
CC to a C-terminal portion of the LZ variant (see also W13955) of  
CC GCN4 and, in some cases, the C-terminal portion of wild-type  
CC p53. The chimeric proteins have DNA binding activity and can  
CC replace lost or insufficient p53 function, providing the means for  
CC pharmacological rescue of p53 function in cancer patients. Nucleic  
CC acids coding for modified p53 constructs can be used for cancer  
CC gene therapy.  
SQ Sequence 359 AA;  
Query Match 100.0%; Score 64; DB 1; Length 359;  
Best Local Similarity 100.0%; Pred. No. 3.07e+00;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Db 249 RPILTIITL 257  
|||||  
QY 1 RPILTIITL 9  
RESULT 7  
ID W13961 standard; Protein; 361 AA.  
AC W13961;  
DE 23-JUN-1997 (first entry)  
DE Chimeric p53 protein.  
KW p53; tumour suppressor; cancer; therapy; cell proliferation;  
KW apoptosis; protein engineering; GCN4; DNA binding.  
OS Chimeric Homo sapiens;  
OS Chimeric synthetic.  
FH Key Location/Qualifiers  
FT region 1..323  
FT /label= p53wt  
FT /note= "amino acids 1-323 of wild-type p53"  
FT region 324..329  
FT /label= Linker  
FT region 330..361  
FT /label= GCN4  
FT /note= "amino acids 250-281 of GCN4 LZ variant"  
PN WO9710843-A1.  
PD 27-MAR-1997.  
PF 20-SEP-1996; U15188.  
PR 22-SEP-1995; US-004802.  
PR 21-AUG-1996; US-697221.  
PA (WIST-) WISTAR INST ANATOMY & BIOLOGY.  
PI Halazonetis TD;  
DR WPI; 97-202618/18.  
PT R284K modified p53 protein having DNA binding ability - useful in  
PT treatment of cancer  
PS Disclosure; Refer to Page 8; 82pp; English.  
CC Chimeric p53 constructs (W13956-67) comprise N-terminal portions  
CC of human wild-type p53 tumour suppressor (see also W13948) linked  
CC to a C-terminal portion of the LZ variant (see also W13955) of  
CC GCN4 and, in some cases, the C-terminal portion of wild-type  
CC p53. The chimeric proteins have DNA binding activity and can  
CC replace lost or insufficient p53 function, providing the means for  
CC pharmacological rescue of p53 function in cancer patients. Nucleic  
CC acids coding for modified p53 constructs can be used for cancer  
CC gene therapy.  
SQ Sequence 361 AA;  
Query Match 100.0%; Score 64; DB 1; Length 361;  
Best Local Similarity 100.0%; Pred. No. 3.07e+00;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Db 249 RPILTIITL 257  
|||||  
QY 1 RPILTIITL 9  
RESULT 8  
ID W13958 standard; Protein; 361 AA.  
AC W13958;

```

DE 25-JUN-1997 (first entry)
KW Chimeric p53 protein.
KW p53; tumour suppressor; cancer; therapy; cell proliferation;
KW apoptosis; protein engineering; GCN4; DNA binding.
OS Chimeric Homo sapiens;
OS Chimeric synthetic.
FH Key Location/Qualifiers
FT region 1..325
FT /label= p53wt
FT /note= "amino acids 1-325 of wild-type p53"
FT region 326..328
FT /label= Linker
FT region 329..361
FT /label= GCN4
FT /note= "amino acids 249-281 of GCN4 LZ variant"
PN W09710843-A1.
PD 27-MAR-1997.
PF 20-SEP-1996; U15188.
PR 22-SEP-1995; US-004802.
PR 21-AUG-1996; US-697221.
PA (WIST-) WISTAR INST ANATOMY & BIOLOGY.
PI Halazonetis TD;
DR WPI; 97-202618/18.
PT R284K modified p53 protein having DNA binding ability - useful in
PT treatment of cancer
PS Disclosure; Refer to Page 8; 82pp; English.
CC Chimeric p53 constructs (W13956-57) comprise N-terminal portions
CC of human wild-type p53 tumour suppressor (see also W13948) linked
CC to a C-terminal portion of the LZ variant (see also W13955) of
CC GCN4 and, in some cases, the C-terminal portion of wild-type
CC p53. The chimeric proteins have DNA binding activity and can
CC replace lost or insufficient p53 function, providing the means for
CC pharmacological rescue of p53 function in cancer patients. Nucleic
CC acids coding for modified p53 constructs can be used for cancer
CC gene therapy.
SQ Sequence 361 AA;

Query Match 100.0%; Score 64; DB 1; Length 361;
Best Local Similarity 100.0%; Pred. No. 3.07e+00;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 249 RPILTIITL 257
QY 1 RPILTIITL 9

RESULT 9
ID W28479 standard; Protein; 363 AA.
AC W28479;
DE Human p53 protein variant V-325 encoded by pEC114.
KW Leucine zipper domain; LZD; oligomerisation domain; mutant; muten;
KW substitution; replacement; transactivation; viral protein VP16; HSV;
KW anti-oncogene; hyperproliferation; cancer; restenosis;
KW tumour suppression; apoptosis.
OS Chimeric - Homo sapiens.
OS Chimeric - Herpes simplex virus.
OS Synthetic.
PN W09704092-A1.
PD 06-FEB-1997.
PF 17-JUL-1996; F01111.
PR 19-JUL-1995; FR-008729.
PA (RHON ) RHONE POULENC RORER SA.
PI Bracco L, Conseiller E;
DR WPI; 97-132633/12.
PT New p53 variants e.g. with oligomerisation domain replaced by
PT leucine zipper - useful for treating hyper-proliferative disorders,
PT esp. cancer and restenosis
PT Claim 30; Page -; 133pp; French.
PS Claimed variants of protein p53 have at least part of the
PS oligomerisation domain deleted and replaced by a leucine zipper
PS domain. The mutants preferably also have at least part of the p53
PS transactivation domain (amino acids 1-74) deleted and replaced by
PS the transactivating domain (TD) from herpes simplex virus viral
PS protein VP16 (amino acids 411-490). The present sequence is that of
PS a specifically claimed p53 variant designated V-325H and comprising
PS the VP16 TD, amino acids 75-325 of human wild-type p53 (but with
PS Arg182 replaced by His) and a leucine zipper domain at the C-terminal.
PS The p53 variants are more active and more stable tumour suppressors
PS and apoptosis-inducing agents than wild-type p53 and are active where
PS the wild-type protein is not, i.e. they are not inactivated by dominant
PS negative or oncogenic mutants, nor by other cellular proteins (because
PS the leucine zipper domain prevents formation of inactive mixed
PS oligomers).
CC (Note: this sequence does not appear in the specification and has
CC been produced by modifying the given sequence of variant V-325).
SQ Sequence 363 AA;

Query Match 100.0%; Score 64; DB 1; Length 363;
Best Local Similarity 100.0%; Pred. No. 3.07e+00;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 257 RPILTIITL 265
QY 1 RPILTIITL 9

RESULT 10
ID W28480 standard; Protein; 363 AA.
AC W28480;
DE Human p53 protein variant V-325H.
KW Leucine zipper domain; LZD; oligomerisation domain; mutant; muten;
KW substitution; replacement; transactivation; viral protein VP16; HSV;
KW anti-oncogene; hyperproliferation; cancer; restenosis;
KW tumour suppression; apoptosis.
OS Chimeric - Homo sapiens.
OS Chimeric - Herpes simplex virus.
OS Synthetic.
PN Key Location/Qualifiers
FT misc_difference 189
FT /note= "Arg residue at position 182 of wild-type
FT p53 has been mutated to His"
FT W09704092-A1.
PD 06-FEB-1997.
PF 17-JUL-1996; F01111.
PR 19-JUL-1995; FR-008729.
PA (RHON ) RHONE POULENC RORER SA.
PI Bracco L, Conseiller E;
DR WPI; 97-132633/12.
PT New p53 variants e.g. with oligomerisation domain replaced by
PT leucine zipper - useful for treating hyper-proliferative disorders,
PT esp. cancer and restenosis
PT Claim 30; Page -; 133pp; French.
PS Claimed variants of protein p53 have at least part of the
PS oligomerisation domain deleted and replaced by a leucine zipper
PS domain. The mutants preferably also have at least part of the p53
PS transactivation domain (amino acids 1-74) deleted and replaced by
PS the transactivating domain (TD) from herpes simplex virus viral
PS protein VP16 (amino acids 411-490). The present sequence is that of
PS a specifically claimed p53 variant designated V-325H and comprising
PS the VP16 TD, amino acids 75-325 of human wild-type p53 (but with
PS Arg182 replaced by His) and a leucine zipper domain at the C-terminal.
PS The p53 variants are more active and more stable tumour suppressors
PS and apoptosis-inducing agents than wild-type p53 and are active where
PS the wild-type protein is not, i.e. they are not inactivated by dominant
PS negative or oncogenic mutants, nor by other cellular proteins (because
PS the leucine zipper domain prevents formation of inactive mixed
PS oligomers).
CC (Note: this sequence does not appear in the specification and has
CC been produced by modifying the given sequence of variant V-325).
SQ Sequence 363 AA;
```



QY 1 RPILTIITL 9  
|||||

RESULT 11  
ID W13975 standard; Protein; 363 AA.  
AC W13975;  
DT 25-JUN-1997 (first entry)  
DE Modified p53 variant p53H273R284del364-393.  
KW p53; tumour suppressor; cancer; therapy; cell proliferation;  
KW apoptosis; protein engineering; DNA binding.  
OS Synthetic.  
PN W09710843-Al.  
PD 27-MAR-1997.  
PF 20-SEP-1996; U15188.  
PR 22-SEP-1995; US-004802.  
PR 21-AUG-1996; US-697221.  
PA (WIST-) WISTAR INST ANATOMY & BIOLOGY.  
PI Halazonetis TD;  
DR WPI; 97-202618/18.  
PT R284K modified p53 protein having DNA binding ability - useful in  
PT treatment of cancer  
PS Example 1: 58-59: 82pp; English.  
CC Modified p53 variant p53H273R284del364-393 (W13975) has the tumour-  
CC derived His273 mutation (see also W13952), a Thr284 to Arg substn.  
CC (see also W13949) and a deletion of the 30 C-terminal amino acids  
CC of wild-type p53 (W13948). His273 is a Class I p53 tumour mutation  
CC that affects DNA binding. The T284R substitution, introduced by  
CC site-directed mutagenesis of p53 DNA, provides a novel p53-DNA  
CC contact between a phosphate of the DNA backbone and p53, and  
CC restores DNA binding. The C-terminal deletion permits in vitro  
CC DNA binding. The construct provides the means for pharmacological  
CC rescue of p53 function in cancer patients. Other modified p53  
CC constructs (W13949-50, W13953-54, W13968-77) have also been  
CC produced. Nucleic acids coding for modified p53 can be used for  
CC cancer gene therapy.  
SQ Sequence 363 AA;

Query Match 100.0%; Score 64; DB 1; Length 363;  
Best Local Similarity 100.0%; Pred. No. 3.07e+00;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 249 RPILTIITL 257  
QY 1 RPILTIITL 9  
|||||

RESULT 12  
ID W13959 standard; Protein; 363 AA.  
AC W13959;  
DT 25-JUN-1997 (first entry)  
DE Chimeric p53 protein.  
KW p53; tumour suppressor; cancer; therapy; cell proliferation;  
KW apoptosis; protein engineering; GCN4; DNA binding.  
OS Chimeric Homo sapiens;  
OS Chimeric synthetic.  
FH Key Location/Qualifiers  
FT region 1..325  
FT /label= p53wt  
FT /note= "amino acids 1-325 of wild-type p53"  
FT region 326..331  
FT /label= Linker  
FT region 332..363  
FT /label= GCN4  
FT /note= "amino acids 250-281 of GCN4 LZ variant"  
PN W09710843-Al.  
PD 27-MAR-1997.  
PF 20-SEP-1996; U15188.  
PR 22-SEP-1995; US-004802.  
PR 21-AUG-1996; US-697221.  
PA (WIST-) WISTAR INST ANATOMY & BIOLOGY.  
PI Halazonetis TD;  
DR WPI; 97-202618/18.

Query Match 100.0%; Score 64; DB 1; Length 374;  
Best Local Similarity 100.0%; Pred. No. 3.07e+00;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

PT R284K modified p53 protein having DNA binding ability - useful in  
PT treatment of cancer  
PS Disclosure; Refer to Page 8; 82pp; English.  
CC Chimeric p53 constructs (W13956-67) comprise N-terminal portions  
CC of human wild-type p53 tumour suppressor (see also W13948) linked  
CC to a C-terminal portion of the LZ variant (see also W13955) of  
CC GCN4 and, in some cases, the C-terminal portion of wild-type  
CC p53. The chimeric proteins have DNA binding activity and can  
CC replace lost or insufficient p53 function, providing the means for  
CC pharmacological rescue of p53 function in cancer patients. Nucleic  
CC acids coding for modified p53 constructs can be used for cancer  
CC gene therapy.  
SQ Sequence 363 AA;

Query Match 100.0%; Score 64; DB 1; Length 363;  
Best Local Similarity 100.0%; Pred. No. 3.07e+00;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 249 RPILTIITL 257  
QY 1 RPILTIITL 9  
|||||

RESULT 13  
ID W28482 standard; Protein; 374 AA.  
AC W28482;  
DT 25-NOV-1997 (first entry)  
DE Human p53 protein variant V-336H.  
KW Leucine zipper domain; LZD; oligomerisation domain; mutant; muten;  
KW substitution; replacement; hyperproliferation; cancer; restenosis;  
KW anti-oncogene; hyperproliferation; cancer; restenosis;  
KW tumour suppression; apoptosis.  
OS Chimeric - Homo sapiens.  
OS Chimeric - Herpes simplex virus.  
OS Synthetic.  
FH Key Location/Qualifiers  
FT misc\_difference 189  
FT /note= "Arg residue at position 182 of wild-type  
FT p53 has been mutated to His"

WO9704092-Al.  
PD 06-FEB-1997.  
PF 17-JUL-1996; F01111.  
PR 19-JUL-1995; FR-008729.  
PA (RHON ) RHONE POULENC RORER SA.  
PI Bracco L, Conseiller E;  
DR WPI; 97-132633/12.  
PT New p53 variants e.g. with oligomerisation domain replaced by  
PT leucine zipper - useful for treating hyper-proliferative disorders,  
PT esp. cancer and restenosis  
PS Claim 31; Page -; 133pp; French.  
CC Claimed variants of protein p53 have at least part of the  
CC oligomerisation domain deleted and replaced by a leucine zipper  
CC domain. The mutants preferably also have at least part of the p53  
CC transactivation domain (amino acids 1-74) deleted and replaced by  
CC the transactivating domain (TD) from herpes simplex virus viral  
CC protein VP16 (amino acids 411-490). The present sequence is that of  
CC a specifically claimed p53 variant designated V-336H and comprising  
CC the VP16 TD, amino acids 75-336 of human wild-type p53 (but with  
CC Arg182 replaced by His) and a leucine zipper domain at the C-terminal.  
CC The p53 variants are more active and more stable tumour suppressors  
CC and apoptosis-inducing agents than wild-type p53 and are active where  
CC the wild-type protein is not, i.e. they are not inactivated by dominant  
CC negative or oncogenic mutants, nor by other cellular proteins (because  
CC the leucine zipper domain prevents formation of inactive mixed  
CC oligomers).  
CC (Note: this sequence does not appear in the specification and has  
CC been produced by modifying the given sequence of variant V-336).  
SQ Sequence 374 AA;

Query Match 100.0%; Score 64; DB 1; Length 374;  
Best Local Similarity 100.0%; Pred. No. 3.07e+00;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Db 257 RPILTIITL 265
|||||
QY 1 RPILTIITL 9

RESULT 14
ID W28481 standard; Protein: 374 AA.
AC W28481;
DT 25-NOV-1997 (first entry)
DE Human p53 protein variant V-336 encoded by p53.
KW Leucine zipper domain; LZD; oligomerisation domain; mutant; mutein;
KW substitution; replacement; transactivation; viral protein VP16; HSV;
KW anti-oncogene; hyperproliferation; cancer; restenosis;
KW tumour suppression; apoptosis.
OS Chimeric - Homo sapiens.
OS Chimeric - Herpes simplex virus.
OS Synthetic.
PN WO9704092-A1.
PD 06-FEB-1997.
PF 17-JUL-1996; F01111.
PR 19-JUL-1995; FR-008729.
PA (RHON ) RHONE POULENC RORER SA.
PI Bracco L, Conseiller E;
DR N-PSDB; T86216.
PT New p53 variants e.g. with oligomerisation domain replaced by
PT leucine zipper - useful for treating hyper-proliferative disorders,
PT esp. cancer and restenosis
PS Claim 31; Pages 78-80; 133pp; French.
CC Claimed variants of protein p53 have at least part of the
CC oligomerisation domain deleted and replaced by a leucine zipper
CC domain. The mutants preferably also have at least part of the p53
CC transactivation domain (amino acids 1-74) deleted and replaced by
CC the transactivating domain (TD) from herpes simplex virus viral
CC protein VP16 (amino acids 411-490). The present sequence is that of
CC a specifically claimed p53 variant designated V-336 and comprising
CC the VP16 TD, amino acids 75-336 of human wild-type p53 and a
CC leucine zipper domain at the C-terminal. The p53 variants are
CC more active and more stable tumour suppressors and apoptosis-inducing
CC agents than wild-type p53 and are active where the wild-type protein
CC is not, i.e. they are not inactivated by dominant negative or oncogenic
CC mutants, nor by other cellular proteins (because the leucine zipper
CC domain prevents formation of inactive mixed oligomers).
SQ Sequence 374 AA;

Query Match 100.0%; Score 64; DB 1; Length 374;
Best Local Similarity 100.0%; Pred. No. 3.07e+00;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 257 RPILTIITL 265
|||||
QY 1 RPILTIITL 9

Search completed: Sat Apr 15 00:43:15 2000
Job time : 37 secs.

DR N-PSDB; T86220.
PT New p53 variants e.g. with oligomerisation domain replaced by
PT leucine zipper - useful for treating hyper-proliferative disorders,
PT esp. cancer and restenosis
PS Claim 35; Pages 85-87; 133pp; French.
CC Claimed variants of protein p53 have at least part of the
CC oligomerisation domain deleted and replaced by a leucine zipper
CC domain. The mutants preferably also have at least part of the p53
CC transactivation domain (amino acids 1-74) deleted and replaced by
CC the transactivating domain (TD) from herpes simplex virus viral
CC protein VP16 (amino acids 411-490). The present sequence is that of
CC a specifically claimed p53 variant designated V-343 and comprising
CC the VP16 TD, amino acids 75-343 of human wild-type p53 and a
CC leucine zipper domain at the C-terminal. The p53 variants are
CC more active and more stable tumour suppressors and apoptosis-inducing
CC agents than wild-type p53 and are active where the wild-type protein
CC is not, i.e. they are not inactivated by dominant negative or oncogenic
CC mutants, nor by other cellular proteins (because the leucine zipper
CC domain prevents formation of inactive mixed oligomers).
SQ Sequence 381 AA;

Query Match 100.0%; Score 64; DB 1; Length 381;
Best Local Similarity 100.0%; Pred. No. 3.07e+00;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 257 RPILTIITL 265
|||||
QY 1 RPILTIITL 9

Search completed: Sat Apr 15 00:43:15 2000
Job time : 37 secs.

Db 257 RPILTIITL 265
|||||
QY 1 RPILTIITL 9

RESULT 15
ID W28489 standard; Protein: 381 AA.
AC W28489;
DT 25-NOV-1997 (first entry)
DE Human p53 protein variant V-343 encoded by p53.
KW Leucine zipper domain; LZD; oligomerisation domain; mutant; mutein;
KW substitution; replacement; transactivation; viral protein VP16; HSV;
KW anti-oncogene; hyperproliferation; cancer; restenosis;
KW tumour suppression; apoptosis.
OS Chimeric - Homo sapiens.
OS Chimeric - Herpes simplex virus.
OS Synthetic.
PN WO9704092-A1.
PD 06-FEB-1997.
PF 17-JUL-1996; F01111.
PR 19-JUL-1995; FR-008729.
PA (RHON ) RHONE POULENC RORER SA.
PI Bracco L, Conseiller E;
DR N-PSDB; T86216.
PT New p53 variants e.g. with oligomerisation domain replaced by
PT leucine zipper - useful for treating hyper-proliferative disorders,
PT esp. cancer and restenosis
PS Claim 31; Pages 78-80; 133pp; French.
CC Claimed variants of protein p53 have at least part of the
CC oligomerisation domain deleted and replaced by a leucine zipper
CC domain. The mutants preferably also have at least part of the p53
CC transactivation domain (amino acids 1-74) deleted and replaced by
CC the transactivating domain (TD) from herpes simplex virus viral
CC protein VP16 (amino acids 411-490). The present sequence is that of
CC a specifically claimed p53 variant designated V-336 and comprising
CC the VP16 TD, amino acids 75-336 of human wild-type p53 and a
CC leucine zipper domain at the C-terminal. The p53 variants are
CC more active and more stable tumour suppressors and apoptosis-inducing
CC agents than wild-type p53 and are active where the wild-type protein
CC is not, i.e. they are not inactivated by dominant negative or oncogenic
CC mutants, nor by other cellular proteins (because the leucine zipper
CC domain prevents formation of inactive mixed oligomers).
SQ Sequence 374 AA;

Query Match 100.0%; Score 64; DB 1; Length 374;
Best Local Similarity 100.0%; Pred. No. 3.07e+00;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 257 RPILTIITL 265
|||||
QY 1 RPILTIITL 9

RESULT 15
ID W28489 standard; Protein: 381 AA.
AC W28489;
DT 25-NOV-1997 (first entry)
DE Human p53 protein variant V-343 encoded by p53.
KW Leucine zipper domain; LZD; oligomerisation domain; mutant; mutein;
KW substitution; replacement; transactivation; viral protein VP16; HSV;
KW anti-oncogene; hyperproliferation; cancer; restenosis;
KW tumour suppression; apoptosis.
OS Chimeric - Homo sapiens.
OS Chimeric - Herpes simplex virus.
OS Synthetic.
PN WO9704092-A1.
PD 06-FEB-1997.
PF 17-JUL-1996; F01111.
PR 19-JUL-1995; FR-008729.
PA (RHON ) RHONE POULENC RORER SA.
PI Bracco L, Conseiller E;
DR N-PSDB; T86216.
PT New p53 variants e.g. with oligomerisation domain replaced by
PT leucine zipper - useful for treating hyper-proliferative disorders,
PT esp. cancer and restenosis
PS Claim 31; Pages 78-80; 133pp; French.
CC Claimed variants of protein p53 have at least part of the
CC oligomerisation domain deleted and replaced by a leucine zipper
CC domain. The mutants preferably also have at least part of the p53
CC transactivation domain (amino acids 1-74) deleted and replaced by
CC the transactivating domain (TD) from herpes simplex virus viral
CC protein VP16 (amino acids 411-490). The present sequence is that of
CC a specifically claimed p53 variant designated V-336 and comprising
CC the VP16 TD, amino acids 75-336 of human wild-type p53 and a
CC leucine zipper domain at the C-terminal. The p53 variants are
CC more active and more stable tumour suppressors and apoptosis-inducing
CC agents than wild-type p53 and are active where the wild-type protein
CC is not, i.e. they are not inactivated by dominant negative or oncogenic
CC mutants, nor by other cellular proteins (because the leucine zipper
CC domain prevents formation of inactive mixed oligomers).
SQ Sequence 374 AA;
```

\*\*\*\*\*  
WIREH  
\*\*\*\*\*  
(TM)  
\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd  
MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Sat Apr 15 00:43:33 2000; MasPar time 3.21 Seconds  
Tabular output not generated.  
112.198 Million cell updates/sec

Title: >US-08-452-843-17  
Description: (1-9) from US08452843.pep  
Perfect Score: 64  
Sequence: 1 RPILITL 9  
Scoring table: PAM 150  
Gap 15  
Searched: 122810 seqs, 40068593 residues  
Post-processing: Minimum Match 0%  
Listing first 45 summaries  
Database: pir62  
1:pir1 2:pir2 3:pir3 4:pir4  
Statistics: Mean 23.550; Variance 29.441; scale 0.800

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES				Pred. No.	
Result No.	Score	Query Match	Length DB ID	Description	
1	64	100.0	363 2	A29376	cellular tumor antigen
2	64	100.0	381 2	S38824	cellular tumor antigen
3	64	100.0	386 2	S51648	cellular tumor antigen
4	64	100.0	390 1	DNMS53	cellular tumor antigen
5	64	100.0	391 2	S02192	cellular tumor antigen
6	64	100.0	391 2	JC6193	tumor suppressor p53
7	64	100.0	393 2	S06594	cellular tumor antigen
8	64	100.0	393 2	JC6176	tumor suppressor prot
9	64	100.0	393 2	DNH053	cellular tumor antigen
10	64	100.0	396 2	JH0633	cellular tumor antigen
11	64	100.0	396 2	JH0631	cellular tumor antigen
12	59	92.2	367 2	S02193	cellular tumor antigen
13	51	79.7	327 2	D71651	octaprenyl-diphosphat
14	50	78.1	328 2	A55215	kdkg 5'-region hypoth
15	50	78.1	522 2	JC4532	cytochrome P450 4F4 p
16	49	76.6	568 2	JQ2206	UL46h protein - Marek
17	49	76.6	626 2	T03547	probable ferrous iron
18	48	75.0	160 2	R44020	TrbH - plasmodium RK2
19	48	75.0	253 2	B69758	conserved hypothetical
20	48	75.0	397 2	G70078	pyrimidine nucleoside
21	48	75.0	533 2	S71617	dimethylaniline monoo
22	48	75.0	533 2	S51131	flavin-containing mon
23	48	75.0	533 2	S71618	dimethylaniline monoo

conserved hypotheticala 3.31e+01  
conserved hypotheticala 3.31e+01  
alpha-1,2-mannosyltra 3.11e+01  
hypothetical protein 5.11e+01  
protein-export membra 5.11e+01  
ferrichrome ABC trans 5.11e+01  
branched-chain amino 5.11e+01  
general sporulation p 5.11e+01  
fasciclin IV precursu 5.11e+01  
probable olfactory re 7.82e+01  
odorant receptor (clo 7.82e+01  
orotate phosphoribosy 7.82e+01  
orotate phosphoribosy 7.82e+01  
cytochrome a3 quinol 7.82e+01  
C02D5.2 protein - Cae 7.82e+01  
cytochrome ba(3) chai 7.82e+01  
celB protein - Escher 7.82e+01  
hypothetical protein 7.82e+01  
probable membrane pro 7.82e+01  
ABC transporter (ATP- 7.82e+01  
glycoprotein B - huma 7.82e+01

ALIGNMENTS

RESULT 1  
ENTRY A29376 #type complete  
TITLE cellular tumor antigen p53 - African clawed frog  
ORGANISM #formal\_name Xenopus laevis #common\_name African clawed frog  
DATE 31-Mar-1989 #sequence\_revision 31-Mar-1989 #text\_change 08-Sep-1997  
ACCESSIONS A29376; S61531; S72313; I51639  
REFERENCE A29376  
#authors Soussi, T.; de Fromental, C.C.; Mechali, M.; May, P.; Kress, M.  
#journal Oncogene (1987) 1:71-78  
#title Cloning and characterization of a cDNA from Xenopus laevis coding for a protein homologous to human and murine p53.  
#cross-references MUID:88143684  
#accession A29376  
#molecule\_type mRNA  
#residues 1-363 #label SOU  
#cross-references EMBL:X05191; NID:964961; PID:964962  
REFERENCE I51639  
#authors Hoever, M.; Clement, J.H.; Wedlich, D.; Montenarh, M.; Knoechel, W.  
#journal Oncogene (1994) 9:109-120  
#title Overexpression of wild-type p53 interferes with normal development in Xenopus laevis embryos.  
#cross-references MUID:94134403  
#accession S61531  
#molecule\_type mRNA  
#residues 1-293,295-363 #label HOE  
#cross-references EMBL:X77546; NID:9468513; PID:9468514  
REFERENCE S72313  
#authors Hoever, M.; Clement, J.; Wedlich, D.; Montenarh, M.; Knoechel, W.  
#submission submitted to the EMBL Data Library, March 1994  
#accession S72313  
#molecule\_type mRNA  
#residues 1-51, 'S', 53-70, 72-293, 295-363 #label HOW  
#cross-references EMBL:X77546; NID:9468513; PID:9468514  
GENETICS p53  
#gene superfamily cellular tumor antigen p53  
CLASSIFICATION apoptosis; cell division control; DNA binding; homotetramer; nucleus; phosphoprotein; transcription regulation; tumor suppressor; zinc  
KEYWORDS  
#binding\_site zinc (Cys, His, Cys, Cys) #status predicted  
#binding\_site phosphoryl-RNA (Ser) (covalent) #status predicted  
FEATURE 150,153,213,217  
362

```

SUMMARY      #length 363 #molecular-weight 40692 #checksum 6648
              predicted
Query Match   100.0%; Score 64; DB 2; Length 363;
Best Local Similarity 100.0%; Pred. No. 8.38e-03;
Matches      9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 224 RPILTIITL 232
| | | | | | | |
Qy 1 RPILTIITL 9

RESULT 2
ENTRY   S38824 #type complete
TITLE   cellular tumor antigen p53, minor splice form - mouse
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE    13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change
17-Mar-1999
ACCESSIONS S38824; S35478
REFERENCE  Arai, N.; Nomura, D.; Yokota, K.; Wolf, D.; Brill, E.;
           Shohat, O.; Rotter, V.
           Mol. Cell. Biol. (1996) 6:3232-3239
           Immunologically distinct p53 molecules generated by
           alternative splicing.
#cross-references EMBL:87064540
#accession S38824
#molecule_type mRNA
#residues 1-381 #label ARA
#cross-references GB:M13874; NID:g200202; PID:g200203
S35478
#authors Han, K.A.; Kulesz-Martin, M.F.
#journal Nucleic Acids Res. (1992) 20:1979-1981
#title Alternatively spliced p53 RNA in transformed and normal cells
of different tissue types.
#cross-references MUID:92253421
#accession S35478
#status nucleic acid sequence not shown; translation not shown
#residues 1-381 #label HAN
#cross-references EMBL:M13874; NID:g200202; PID:g200203
#note the nucleotide sequence was submitted to the EMBL Data
Library, July 1988
COMMENT This sequence, produced by alternative splicing of the tenth
intron, lacks the carboxyl-terminal sequence necessary for
covalent attachment of RNA. The function of this minor splice
form is not known.
CLASSIFICATION #superfamily cellular tumor antigen p53
KEYWORDS alternative splicing; phosphoprotein; zinc
FEATURE
1-44 #domain transcription activation #status predicted
#label TRA\
16-26 #region conserved region I\
93-289 #domain DNA-binding core #status predicted #label DBC\
108-121 #region L1 loop\
114-139 #region conserved region II\
160-192 #region L2 loop\
168-178 #region conserved region III\
231-252 #region conserved region IV\
233-248 #region L3 loop\
267-283 #region conserved region V\
313-319 #region nuclear location signal\
319-357 #region tetramer association\
7,9,12,18,23,37 #binding_site phosphate (Ser) (covalent) #status
predicted\
173,176,235,239 #binding_site zinc (Cys, His, Cys, Cys) #status
predicted\
312 #binding_site phosphate (Ser) (covalent) (by cdc2
kinase) #status predicted
SUMMARY #length 381 #molecular-weight 42498 #checksum 8703
Query Match 100.0%; Score 64; DB 2; Length 381;
Best Local Similarity 100.0%; Pred. No. 8.38e-03;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 246 RPILTIITL 254
| | | | | | | |
Qy 1 RPILTIITL 9

RESULT 3
ENTRY   S51648 #type complete
TITLE   cellular tumor antigen p53 - bovine
ORGANISM #formal_name Bos primigenius taurus #common_name cattle
DATE    07-May-1995 #sequence_revision 01-Sep-1995 #text_change
08-Sep-1997
ACCESSIONS S51648
REFERENCE  Dequiedt, F.; Willems, L.; Burny, A.; Kettmann, R.
           Submitted to the EMBL Data Library, September 1994
           Nucleotide sequence of the ovine p53 tumor-suppressor gene
           cDNA and its genomic organisation.
#accession S51648
#status preliminary
#molecule_type mRNA
#residues 1-386 #label DEQ
#cross-references EMBL:X81704; NID:g602332; PID:g602333
CLASSIFICATION #superfamily cellular tumor antigen p53
KEYWORDS apoptosis; cell division control; DNA binding; homotetramer;
phosphoprotein; transcription regulation; tumor suppressor;
zinc
FEATURE
168,171,231,235 #binding_site zinc (Cys, His, Cys, Cys) #status
predicted\
385 #binding_site phosphoryl-RNA (Ser) (covalent) #status
predicted
SUMMARY #length 386 #molecular-weight 43255 #checksum 7025
Query Match 100.0%; Score 64; DB 2; Length 386;
Best Local Similarity 100.0%; Pred. No. 8.38e-03;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 242 RPILTIITL 250
| | | | | | | |
Qy 1 RPILTIITL 9

RESULT 4
ENTRY   DNMS53 #type complete
TITLE   cellular tumor antigen p53 - mouse
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE    28-Aug-1985 #sequence_revision 04-Oct-1996 #text_change
19-Feb-1999
ACCESSIONS A22739; S06336; A02684; S38822; S38823; S40014; I48703
REFERENCE  A22739
           authors Biernz, B.; Zakut-Houri, R.; Givol, D.; Oren, M.
           #journal EMBO J. (1984) 3:2179-2183
           #cross-references MUID:85027173
           #accession A22739
           #molecule_type DNA
           #residues 1-134, 'V', 136-390 #label BIE
           #cross-references GB:X00876; NID:g871420; PID:g871421; GB:X01237;
           GB:K01700; NID:g53575; PID:g53576
           S06336
           authors Chumakov, P.M.
           #journal Bioorg. Khim. (1987) 13:1691-1694
           #title Primary structure of DNA complementary to murine oncoprotein
           p53 mRNA.
           #cross-references MUID:88221682
           #accession S06336
           #status not compared with conceptual translation
           #molecule_type mRNA
           #residues 1-134, 'V', 136-390 #label CHU
           A02684
           REFERENCE

```

```

#authors      Zakut-Houri, R.; Oren, M.; Bienz, B.; Lavie, V.; Hazum, S.;
               Givol, D.
#journal      Nature (1983) 306:594-597
#title       A single gene and a pseudogene for the cellular tumour
               antigen p53.
#cross-references MUID:84068204
#accession    A02684
##molecule_type mRNA
##residues    1-159, 'H', '161-167', 'G', '169-233', 'I', '235-390' #label ZAK
##cross-references GB:X01237; GB:X01700; NID:G53575
REFERENCE
#authors      Arai, N.; Nomura, D.; Yokota, K.; Wolf, D.; Brill, E.;
               Shohat, O.; Rotter, V.
#journal      Mol. Cell. Biol. (1986) 6:3232-3239
#title       Immunologically distinct p53 molecules generated by
               alternative splicing.
#cross-references MUID:87064640
#accession    S38822
##status      preliminary
##molecule_type mRNA
##residues    1-390 #label ARA1
##cross-references EMBL:M13872; NID:G200199
#accession    S38823
##status      preliminary
##molecule_type mRNA
##residues    1-167, 'G', '169-233', 'I', '235-390' #label ARA2
##cross-references EMBL:M13873
REFERENCE
#authors      Arai, N.; Nomura, D.; Yokota, K.; Wolf, D.; Brill, E.;
               Shohat, O.; Rotter, V.
#submission   submitted to the EMBL Data Library, July 1988
#accession    S40014
##molecule_type mRNA
##residues    1-167, 'G', '169-390' #label ARA3
##cross-references EMBL:M13873; NID:G200200; PID:G200201
REFERENCE
#authors      Jenkins, J.R.; Rudge, K.; Redmond, S.; Wade-Evans, A.
#journal      Nucleic Acids Res. (1984) 12:5609-5626
#title       Cloning and expression analysis of full length mouse cDNA
               sequences encoding the transformation associated protein
               p53.
#cross-references MUID:84272240
#accession    I48703
##status      preliminary; translated from GB/EMBL/DBJ
##molecule_type mRNA
##residues    1-47, 'R', '49-78', 'QW', '82-390' #label RES
##cross-references EMBL:X00741; NID:G53570; PID:G53571
COMMENT      This DNA-binding protein plays an essential role in the regulation
               of cell division, as it is required for the transition from phase
               G0 to G1 of the cell cycle.
COMMENT      The tetramer association region may exhibit a beta-turn,
               beta-sheet, beta-turn, alpha-helix motif.
CLASSIFICATION #superfamily cellular tumor antigen p53
KEYWORDS       apoptosis; cell division control; DNA binding; homotetramer;
               phosphoprotein; transcription regulation; tumor suppressor;
               zinc
FEATURE
1-44          #domain transcription activation #status predicted
16-26         #region conserved region I\
99-289        #domain DNA-binding core #status predicted #label DBC\
108-121        #region L1 loop\
114-139        #region conserved region II\
160-192        #region L2 loop\
168-178        #region conserved region III\
231-252        #region conserved region IV\
267-283        #region conserved region V\
313-319        #region nuclear location signal\
319-357        #region tetramer association\
7,9,12,18,23,37 #binding_site phosphate (Ser) (covalent) #status
               predicted\
173,176,235,239 #binding_site zinc (Cys, His, Cys, Cys) #status

```

```

312           predicted\
               #binding_site phosphate (Ser) (covalent) (by cdc2
               kinase) #status predicted\
389           #binding_site phosphoryl-RNA (Ser) (covalent) #status
               predicted\
SUMMARY       #length 390 #molecular-weight 43458 #checksum 1260
Query Match   100.0%; Score 64; DB 1; Length 390;
Best Local Similarity 100.0%; Pred. No. 8.38e-03;
Matches       9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db           246 RPILTIITL 254
QY           1 RPILTIITL 9
               |||||
RESULT       5
ENTRY       S02192 #type complete
TITLE       cellular tumor antigen p53 - rat
ALTERNATE_NAMES gene p53 protein; nuclear oncoprotein p53
ORGANISM      #formal_name Rattus norvegicus #common_name Norway rat
.DATE        18-Oct-1989 #sequence_revision 18-Oct-1989 #text_change
               17-Mar-1999
ACCESSIONS   S02192; S41149
REFERENCE     Soussi, T.; de Fromental, C.C.; Breugnot, C.; May, E.
               Nucleic Acids Res. (1988) 16:11384
               Nucleotide sequence of a cDNA encoding the rat p53 nuclear
               oncoprotein.
               #cross-references MUID:89083585
               #accession S02192
               ##molecule_type mRNA
               ##residues 1-391 #label SOU
               ##cross-references EMBL:X13058; NID:G56828; PID:G56829
REFERENCE     S41149
               #authors Hulla, J.E.; Schneider, R.P.
               #journal Nucleic Acids Res. (1993) 21:713-717
               #title Structure of the rat p53 tumor suppressor gene.
               #cross-references MUID:93181268
               #accession S41149
               ##status preliminary; nucleic acid sequence not shown;
               translation not shown
               ##molecule_type DNA
               ##residues 1-173, 'W', '175-391' #label HUL
               ##cross-references EMBL:L07909
               ##note the nucleotide sequence was submitted to the EMBL Data
               Library, December 1992
GENETICS
#introns      25/2; 32/3; 123/3; 185/1; 259/2; 305/1; 329/3; 365/2
CLASSIFICATION #superfamily cellular tumor antigen p53
KEYWORDS       apoptosis; cell division control; DNA binding; homotetramer;
               nucleus; phosphoprotein; transcription regulation; tumor
               suppressor; zinc
FEATURE
174,177,236,240 #binding_site zinc (Cys, His, Cys, Cys) #status
               predicted\
390           #binding_site phosphoryl-RNA (Ser) (covalent) #status
               predicted\
SUMMARY       #length 391 #molecular-weight 43451 #checksum 7105
Query Match   100.0%; Score 64; DB 2; Length 391;
Best Local Similarity 100.0%; Pred. No. 8.38e-03;
Matches       9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db           247 RPILTIITL 255
QY           1 RPILTIITL 9
               |||||
RESULT       6
ENTRY       JC6193 #type complete
TITLE       tumor suppressor p53 - rabbit
ORGANISM      #formal_name Oryctolagus cuniculus #common_name domestic

```

DATE	11-Apr-1997	#sequence_revision	09-May-1997	#text_change
ACCESSIONS	JC6193			
REFERENCE	JC6193			
#authors	Le Goas, F.; May, P.; Ronco, P.; de Fromental, C.C.			
#journal	Gene (1997) 185:169-173			
#title	CDNA cloning and immunological characterization of rabbit p53.			
#cross-references	MUID:97208869			
#accession	JC6193			
#molecule_type	mRNA			
#residues	1-391	#label	LEA	
#cross-references	EMBL:X90592; NID:q1532043; PID:e194962; PID:g1532044			
GENETICS				
#gene	p53			
CLASSIFICATION	#superfamily cellular tumor antigen p53			
KEYWORDS	tumor			
SUMMARY	#length 391 #molecular-weight 43435 #checksum 4367			
Query Match	100.0%; Score 64; DB 2: Length 391;			
Best Local Similarity	100.0%; Pred. No. 8.38e-03;			
Matches	9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
Db	246 RPILITIIL 254			
Qy	1 RPILITIIL 9			
RESULT	7			
ENTRY	S06594	#type	complete	
TITLE	cellular tumor antigen p53 - green monkey			
ORGANISM	#formal_name Cercopithecus aethiops #common_name green monkey grivet			
DATE	28-Feb-1990	#sequence_revision	28-Feb-1990	#text_change
ACCESSIONS	S06594			
REFERENCE	S06594			
#authors	Rigaudy, P.; Eckhart, W.			
#journal	Nucleic Acids Res. (1989) 17:8375			
#title	Nucleotide sequence of a cDNA encoding the monkey cellular phosphoprotein p53.			
#cross-references	MUID:90045967			
#accession	S06594			
#molecule_type	mRNA			
#residues	1-393	#label	RIG	
#cross-references	EMBL:X16384; NID:q22795; PID:g22796			
CLASSIFICATION	#superfamily cellular tumor antigen p53			
KEYWORDS	apoptosis; cell division control; DNA binding; homotetramer; nucleus; phosphoprotein; transcription regulation; tumor suppressor; zinc			
FEATURE				
176,179,238,242	#binding_site zinc (Cys, His, Cys, Cys) #status predicted			
392	#binding_site phosphoryl-RNA (Ser) (covalent) #status predicted			
SUMMARY	#length 393 #molecular-weight 43696 #checksum 4263			
Query Match	100.0%; Score 64; DB 2: Length 393;			
Best Local Similarity	100.0%; Pred. No. 8.38e-03;			
Matches	9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
Db	249 RPILITIIL 257			
Qy	1 RPILITIIL 9			
RESULT	8			
ENTRY	JC6176	#type	complete	
TITLE	tumor suppressor protein p53 - Chinese hamster			
ORGANISM	#formal_name Cricetulus griseus #common_name Chinese hamster			
DATE	11-Apr-1997	#sequence_revision	09-May-1997	#text_change
	08-Sep-1997			

#submission submitted to the EMBL Data Library, August 1990  
#accession S40773

##molecule\_type DNA  
##residues 1-393 ##label CHU  
##cross-references EMBL:X54156; NID:g35213; PID:g35214

## REFERENCE

#authors Matlashewski, G.; Lamb, P.; Pim, D.; Peacock, J.; Crawford, L.; Benchimol, S.  
#journal EMBO J. (1984) 3:3257-3262  
#title Isolation and characterization of a human p53 cDNA clone: expression of the human p53 gene.

#cross-references MUID:85126934

#accession S42669

##molecule\_type mRNA

##residues 101-393 ##label MK11

##cross-references EMBL:X01405; NID:g35215; PID:g642241

## REFERENCE

#authors Zakut-Houri, R.; Bienen-Tadmor, B.; Givol, D.; Oren, M.  
#journal EMBO J. (1985) 4:1251-1255  
#title Human p53 cellular tumor antigen: cDNA sequence and expression in COS cells.

#cross-references MUID:85230577

#accession A22837

##molecule\_type mRNA

##residues 1-71, 'P', 73-393 ##label ZAK

##cross-references EMBL:X02469; EMBL:M60950; NID:g35209; PID:g35210

## REFERENCE

#authors Harlow, E.; Williamson, N.M.; Ralston, R.; Helfman, D.M.; Adams, T.E.

#journal Mol. Cell. Biol. (1985) 5:1601-1610

#title Molecular cloning and in vitro expression of a cDNA clone for human cellular tumor antigen p53.

#cross-references MUID:85267676

#accession A55060

##molecule\_type mRNA

##residues 1-71, 'P', 73-272, 'H', 274-393 ##label HAR

##cross-references GB:X0199; NID:g189478; PID:g189479

##experimental\_source clone PR4-2, cell line A431

## REFERENCE

#authors Harris, N.; Brill, E.; Shohat, O.; Prokocimer, M.; Wolf, D.; Arai, N.; Rotter, V.

#journal Mol. Cell. Biol. (1986) 6:4650-4656

#title Molecular basis for heterogeneity of the human p53 protein.

#cross-references MUID:87089826

#accession A25397

##molecule\_type mRNA

##residues 1-78, 'T', 80-393 ##label HAR1

##cross-references EMBL:M14694; NID:g339813; PID:g339814

##experimental\_source clone p53-H-1, transformed hybridoma SV-80 cell line

#accession B25397

##molecule\_type mRNA

##residues 1-71, 'P', 73-78, 'T', 80-393 ##label HAR2

##cross-references EMBL:M14695; NID:g339815; PID:g339816

##experimental\_source clone p53-H-19, transformed hybridoma SV-80 cell line

## REFERENCE

#authors Matlashewski, G.J.; Tuck, S.; Pim, D.; Lamb, P.; Schneider, J.; Crawford, L.V.

#journal Mol. Cell. Biol. (1987) 7:961-963

#title Primary structure polymorphism at amino acid residue 72 of human p53.

#cross-references MUID:87144273

## REFERENCE

#molecule\_type mRNA; DNA

##residues 66-71, 'P', 73-79 ##label MK12

##experimental\_source clone lambda C13

##note 72-Cys was also found, and appears to represent a polymorphism

#accession S42453

##molecule\_type mRNA; DNA

##residues 66-79 ##label MK13

##experimental\_source clone J6K

## REFERENCE

I38082  
#authors Farrell, P.J.; Allan, G.J.; Shanahan, F.; Vousden, K.H.; Crook, T.

#journal EMBO J. (1991) 10:2879-2887

#title p53 is frequently mutated in Burkitt's lymphoma cell lines.

#cross-references MUID:92007731

#accession I38082

##status translated from GB/EMBL/DBDJ

##molecule\_type mRNA

##residues 1-189, 'LLSILSEWKEICVWSIWMETLFDIVWCPMSRLRLALT', 'VPSTTTCTVTPAWAA' ##label F01

##cross-references EMBL:X60010; NID:g506432; PID:g506443

##note deletion of a C nucleotide causes a frameshift at position 566

#accession I38083

##status translated from GB/EMBL/DBDJ

##molecule\_type mRNA

##residues 1-192, 'R', 194-393 ##label F02

##cross-references EMBL:X60011; NID:g506434; PID:g506435

#accession I38084

##status translated from GB/EMBL/DBDJ

##molecule\_type mRNA

##residues 1-393 ##label F03

##cross-references EMBL:X60012; NID:g506436; PID:g506437

#accession I38085

##status translated from GB/EMBL/DBDJ

##molecule\_type mRNA

##residues 1-245, 'T', 247-393 ##label F04

##cross-references EMBL:X60013; NID:g506438; PID:g506439

#accession I38086

##status translated from GB/EMBL/DBDJ

##molecule\_type mRNA

##residues 1-236, 'I', 238-393 ##label F05

##cross-references EMBL:X60014; NID:g506440; PID:g506441

#accession I38087

##status translated from GB/EMBL/DBDJ

##molecule\_type mRNA

##residues 1-247, 'Q', 249-393 ##label F06

##cross-references EMBL:X60015; NID:g506442; PID:g506443

#accession I38088

##status translated from GB/EMBL/DBDJ

##molecule\_type mRNA

##residues 1-71, 'P', 73-237, 'Y', 239-393 ##label F07

##cross-references EMBL:X60016; NID:g506444; PID:g506445

#accession I38089

##status translated from GB/EMBL/DBDJ

##molecule\_type mRNA

##residues 1-247, 'Q', 249-393 ##label F08

##cross-references EMBL:X60017; NID:g506446; PID:g506447

#accession I38090

##status translated from GB/EMBL/DBDJ

##molecule\_type mRNA

##residues 1-71, 'P', 73-162, 'H', 164-393 ##label F09

##cross-references EMBL:X60018; NID:g506448; PID:g506449

#accession I38091

##status translated from GB/EMBL/DBDJ

##molecule\_type mRNA

##residues 1-212, 'Q', 214-393 ##label F10

##cross-references EMBL:X60019; NID:g506450; PID:g506451

#accession I38092

##status translated from GB/EMBL/DBDJ

##molecule\_type mRNA

##residues 1-253, 'D', 255-393 ##label F11

##cross-references EMBL:X60020; NID:g506452; PID:g506453

##note all sequences submitted to the EMBL/GenBank/DBJ databases June 1991

## REFERENCE

I38093

#authors Futreal, P.A.; Barrett, J.C.; Wiseman, R.W.

#journal Nucleic Acids Res. (1991) 19:6977

#title An Alu polymorphism intragenic to the TP53 gene.

#cross-references MUID:92107726

#accession I38093

##status translated from GB/EMBL/DBDJ

```

##molecule_type DNA
##residues 1-393 ##label FUT
##cross-references EMBL:X54156; NID:g35213; PID:g35214
REFERENCE A44905
#authors Yamada, Y.; Yoshida, T.; Hayashi, K.; Sekiya, T.; Yokota, J.;
Hirohashi, S.; Nakatani, K.; Nakano, H.; Sugimura, T.;
Terada, M.
#journal Cancer Res. (1991) 51:5800-5805
#title p53 gene mutations in gastric cancer metastases and in
gastric cancer cell lines derived from metastases.
#cross-references MUID:92034678
#accession A44905
...
Note: remainder of annotations omitted.
Query Match 100.0%; Score 64; DB 1; Length 393;
Best Local Similarity 100.0%; Pred. No. 8.38e-03;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 249 RPILTIITL 257
|||||
QY 1 RPILTIITL 9

RESULT 10
ENTRY JH0633 #type complete
TITLE cellular tumor antigen p53 - golden hamster
ALTERNATE_NAMES tumor-suppressor protein p53
ORGANISM #formal_name Mesocricetus auratus #common_name golden hamster
DATE 17-Aug-1992 #sequence_revision 17-Aug-1992 #text_change
08-Sep-1997
ACCESSIONS JH0633
REFERENCE JH0633
#authors Legros, Y.; McIntyre, P.; Soussi, T.
#journal Gene (1992) 112:247-250
#title The cDNA cloning and immunological characterization of
hamster p53.
#cross-references MUID:92210007
#accession JH0633
##molecule_type mRNA
##residues 1-396 ##label LEG
##cross-references GB:M75144; NID:g191414; PID:g191415
##experimental_source kidney, strain MPI
GENETICS p53
#gene #superfamily cellular tumor antigen p53
CLASSIFICATION apoptosis; cell division control; DNA binding; homotetramer;
KEYWORDS nucleus; phosphoprotein; transcription regulation; tumor
suppressor; zinc
FEATURE
179,182,241,245 #binding_site zinc (Cys, His, Cys, Cys) #status
predicted
395 #binding_site phosphoryl-RNA (Ser) (covalent) #status
predicted
SUMMARY #length 396 #molecular-weight 43631 #checksum 6617

Query Match 100.0%; Score 64; DB 2; Length 396;
Best Local Similarity 100.0%; Pred. No. 8.38e-03;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 252 RPILTIITL 260
|||||
QY 1 RPILTIITL 9

RESULT 11
ENTRY JH0631 #type complete
TITLE cellular tumor antigen p53 - rainbow trout
ORGANISM #formal_name Oncorhynchus mykiss #common_name rainbow trout
DATE 17-Aug-1992 #sequence_revision 17-Aug-1992 #text_change
08-Sep-1997
ACCESSIONS JH0631
REFERENCE JH0631

```

```

#authors de Fromental, C.C.; Pakdel, F.; Chapus, A.; Baney, C.; May,
P.; Soussi, T.
#journal Gene (1992) 112:241-245
#title Rainbow trout p53: cDNA cloning and biochemical
characterization.
#cross-references MUID:92210006
#accession JH0631
##molecule_type mRNA
##residues 1-396 ##label DEF
##cross-references GB:M75145; NID:g213828; PID:g213829
##experimental_source liver
COMMENT This protein is the product of a tumor suppressor gene, p53, whose
inactivation leads to cell transformation or neoplasia.
CLASSIFICATION #superfamily cellular tumor antigen p53
KEYWORDS apoptosis; cell division control; DNA binding; homotetramer;
nucleus; phosphoprotein; transcription regulation; tumor
suppressor; zinc
FEATURE
164,167,227,231 #binding_site zinc (Cys, His, Cys, Cys) #status
predicted
395 #binding_site phosphoryl-RNA (Ser) (covalent) #status
predicted
SUMMARY #length 396 #molecular-weight 43966 #checksum 9018

Query Match 100.0%; Score 64; DB 2; Length 396;
Best Local Similarity 100.0%; Pred. No. 8.38e-03;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 238 RPILTIITL 246
|||||
QY 1 RPILTIITL 9

RESULT 12
ENTRY S02193 #type complete
TITLE cellular tumor antigen p53 - chicken
ALTERNATE_NAMES nuclear oncoprotein p53
ORGANISM #formal_name Gallus gallus #common_name chicken
DATE 18-Oct-1989 #sequence_revision 18-Oct-1989 #text_change
08-Sep-1997
ACCESSIONS S02193
REFERENCE S02193
#authors Soussi, T.; Begue, A.; Kress, M.; Stehelin, D.; May, P.
#journal Nucleic Acids Res. (1988) 16:11383
#title Nucleotide sequence of a cDNA encoding the chicken p53
nuclear oncoprotein.
#cross-references MUID:89083584
#accession S02193
##molecule_type mRNA
##residues 1-367 ##label SOU
##cross-references EMBL:X13057; NID:g63740; PID:g63741
CLASSIFICATION #superfamily cellular tumor antigen p53
KEYWORDS apoptosis; cell division control; DNA binding; homotetramer;
nucleus; phosphoprotein; transcription regulation; tumor
suppressor; zinc
FEATURE
161,164,224,228 #binding_site zinc (Cys, His, Cys, Cys) #status
predicted
366 #binding_site phosphoryl-RNA (Ser) (covalent) #status
predicted
SUMMARY #length 367 #molecular-weight 40169 #checksum 5094

Query Match 92.2%; Score 59; DB 2; Length 367;
Best Local Similarity 88.9%; Pred. No. 1.12e-01;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Db 235 RPILTIITL 243
|||||
QY 1 RPILTIITL 9

RESULT 13
ENTRY D71651 #type complete

```



TITLE octaprenyl-diphosphate synthase (ispB) RP479 - Rickettsia prowazekii

ORGANISM #formal\_name Rickettsia prowazekii

DATE 21-Nov-1998 #sequence\_revision 21-Nov-1998 #text\_change 21-Nov-1998

ACCESSIONS D71651

REFERENCE A71630

#authors Andersson, S.G.E.; Zomorodipour, A.; Andersson, J.O.; Sacheritz-Ponten, T.; Alsmark, U.C.M.; Podowski, R.M.; Naeslund, A.K.; Eriksson, A.S.; Winkler, H.H.; Kurland, C.G.

#journal Nature (1998) 396:133-140

#title The genome sequence of Rickettsia prowazekii and the origin of mitochondria.

#accession D71651

#status preliminary; nucleic acid sequence not shown; translation not shown

##molecule\_type DNA

##residues 1-327 #label AND

##cross-references GB:AJ235272; GB:AJ235269; NID:g3861033; PID:el342778; PID:g3861034

##experimental\_source strain Madrid E

GENETICS

#gene ispB; RP479

SUMMARY #length 327 #molecular-weight 37187 #checksum 1234

Query Match 79.7%; Score 51; DB 2; Length 327;

Best Local Similarity 87.5%; Pred. No. 5.47e+00;

Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 49 RP1LTIIT 56

||:|||||

QY 1 RP1LTIIT 8

RESULT 14

ENTRY A55215 #type complete

TITLE KdGK 5'-region hypothetical protein 1 - Erwinia chrysanthemi

ORGANISM #formal\_name Erwinia chrysanthemi

DATE 05-May-1995 #sequence\_revision 05-May-1995 #text\_change 09-Sep-1997

ACCESSIONS A55215

REFERENCE A55215

#authors Hugouvieux-Cotte-Pattat, N.; Nasser, W.; Robert-Baudouy, J.

#journal J. Bacteriol. (1994) 176:2386-2392

#title Molecular characterization of the Erwinia chrysanthemi KdGK gene involved in pectin degradation.

#cross-references MUID:94209241

#accession A55215 preliminary

##status preliminary

##molecule\_type DNA

##residues 1-328 #label HUG

##cross-references GB:X75047; NID:g495246; PID:g495247

SUMMARY #length 328 #molecular-weight 35998 #checksum 125

Query Match 78.1%; Score 50; DB 2; Length 328;

Best Local Similarity 55.6%; Pred. No. 8.66e+00;

Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 216 RPVMTLIAL 224

||:|||||

QY 1 RP1LTIITL 9

RESULT 15

ENTRY JC4532 #type complete

TITLE cytochrome P450 4F4 protein - rat

ORGANISM #formal\_name Rattus norvegicus #common\_name Norway rat

DATE 15-Feb-1996 #sequence\_revision 19-Apr-1996 #text\_change 05-Mar-1999

ACCESSIONS JC4532

REFERENCE JC4532

#authors Kawashima, H.; Strobel, H.W.

#journal Biochem. Biophys. Res. Commun. (1995) 217:1137-1144

#title cDNA cloning of three new forms of rat brain cytochrome P450 belonging to the CYP4F subfamily.

#cross-references MUID:96125358

#accession JC4532

##molecule\_type mRNA

##residues 1-522 #label KAW

##cross-references GB:U39206; NID:g1146435; PID:g1146436

##experimental\_source brain

CLASSIFICATION #superfamily human cytochrome P450 CYP4B1; cytochrome P450 homology

KEYWORDS brain; chromoprotein; heme; iron

FEATURE 468 #binding\_site heme iron (Cys) (axial ligand) #status predicted

SUMMARY #length 522 #molecular-weight 60049 #checksum 177

Query Match 78.1%; Score 50; DB 2; Length 522;

Best Local Similarity 87.3%; Pred. No. 8.66e+00;

Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 94 PILPIITL 101

||:|||||

QY 2 PILTIITL 9

Search completed: Sat Apr 15 00:43:50 2000

Job time : 17 secs.

**THIS PAGE BLANK (USPTO)**

\*\*\*\*\*  
[W][O][R][L][D]  
\*\*\*\*\*  
(TM)  
\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Sat Apr 15 00:44:07 2000; MasPar time 3.10 Seconds  
Tabular output not generated. 86.819 Million cell updates/sec

Title: >US-08-452-843-17  
Description: (1-9) from US08452843.pep  
Perfect Score: 64  
Sequence: 1 RPILTIITL 9

Scoring table: PAM 150  
Gap 15

Searched: 82229 seqs, 29864866 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: swiss-prot38  
1:swissprot

Statistics: Mean 24.215; Variance 25.763; scale 0.940

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	64	100.0	207	1 P53_EQUAS	CELLULAR TUMOR ANTIGEN	9.32e-04
2	64	100.0	280	1 P53_HORSE	CELLULAR TUMOR ANTIGEN	9.32e-04
3	64	100.0	314	1 P53_SPEBE	CELLULAR TUMOR ANTIGEN	9.32e-04
4	64	100.0	363	1 P53_XENLA	CELLULAR TUMOR ANTIGEN	9.32e-04
5	64	100.0	373	1 P53_BRARE	CELLULAR TUMOR ANTIGEN	9.32e-04
6	64	100.0	381	1 P53_CANFA	CELLULAR TUMOR ANTIGEN	9.32e-04
7	64	100.0	382	1 P53_SHEEP	CELLULAR TUMOR ANTIGEN	9.32e-04
8	64	100.0	386	1 P53_BOVIN	CELLULAR TUMOR ANTIGEN	9.32e-04
9	64	100.0	390	1 P53_MOUSE	CELLULAR TUMOR ANTIGEN	9.32e-04
10	64	100.0	391	1 P53_RAT	CELLULAR TUMOR ANTIGEN	9.32e-04
11	64	100.0	391	1 P53_RABIT	CELLULAR TUMOR ANTIGEN	9.32e-04
12	64	100.0	393	1 P53_MACFA	CELLULAR TUMOR ANTIGEN	9.32e-04
13	64	100.0	393	1 P53_HUMAN	CELLULAR TUMOR ANTIGEN	9.32e-04
14	64	100.0	393	1 P53_CRAE	CELLULAR TUMOR ANTIGEN	9.32e-04
15	64	100.0	393	1 P53_MACMU	CELLULAR TUMOR ANTIGEN	9.32e-04
16	64	100.0	393	1 P53_CRIGR	CELLULAR TUMOR ANTIGEN	9.32e-04
17	64	100.0	396	1 P53_SALIR	CELLULAR TUMOR ANTIGEN	9.32e-04
18	64	100.0	396	1 P53_MESAU	CELLULAR TUMOR ANTIGEN	9.32e-04
19	59	92.2	351	1 P53_ORYLA	CELLULAR TUMOR ANTIGEN	1.83e-02
20	59	92.2	367	1 P53_CHICK	CELLULAR TUMOR ANTIGEN	1.83e-02
21	58	90.6	386	1 P53_FELCA	CELLULAR TUMOR ANTIGEN	3.25e-02
22	51	79.7	366	1 P53_PLAFA	CELLULAR TUMOR ANTIGEN	1.56e-00
23	50	78.1	328	1 YHVD_ERWCH	HYPOTHETICAL 36.0 KD P	2.63e+00

24	50	78.1	522	1	CPF4_RAT	CYTOCHROME P450 4F4 (E	2.63e+00
25	48	75.0	397	1	YXJA_BAGSU	HYPOTHETICAL 43.7 KD P	7.31e+00
26	48	75.0	532	1	FM05_CAYPO	DIMETHYLANILINE MONOOX	7.31e+00
27	48	75.0	532	1	FM05_HUMAN	DIMETHYLANILINE MONOOX	7.31e+00
28	47	73.4	464	1	KTR4_YEAST	PROBABLE MANNOSYLTRANS	1.20e+01
29	46	71.9	100	1	VG10_HSVB	HYPOTHETICAL GENE 10 P	1.97e+01
30	46	71.9	311	1	SECF_RICPR	PROTEIN-EXPORT MEMBRAN	1.97e+01
31	46	71.9	439	1	BRNQ_SALTY	BRANCHED-CHAIN AMINO A	1.97e+01
32	46	71.9	439	1	BRNQ_ECOLI	BRANCHED-CHAIN AMINO A	1.97e+01
33	46	71.9	698	1	GSGL_YEAST	SPOULATORY PROTEIN GS	1.97e+01
34	45	70.3	222	1	OL7B_MOUSE	OLFACTORY RECEPTOR 7B	3.18e+01
35	45	70.3	226	1	PYRE_YEAST	OROTATE PHOSPHORIBOSYL	3.18e+01
36	45	70.3	227	1	PYRX_YEAST	OROTATE PHOSPHORIBOSYL	3.18e+01
37	45	70.3	323	1	YKJ2_CAEEL	HYPOTHETICAL 36.9 KD P	3.18e+01
38	45	70.3	323	1	QOX2_BAGSU	QUINOL OXIDASE POLYPEP	3.18e+01
39	45	70.3	452	1	PTCC_ECOLI	PTS SYSTEM, CELLOBIOSE	3.18e+01
40	45	70.3	604	1	YFIC_BAGSU	HYPOTHETICAL ABC TRANS	3.18e+01
41	45	70.3	868	1	VGLB_VZVD	GLYCOPROTEIN B PRECURS	3.18e+01
42	44	68.8	119	1	FRDD_ECOLI	FUMARATE REDUCTASE 13	5.10e+01
43	44	68.8	348	1	NU2M_SQUAC	NADH-UBIQUINONE OXIDOR	5.10e+01
44	44	68.8	431	1	CIT1_ECOLI	CITRATE-PROTON SYMPORT	5.10e+01
45	44	68.8	634	1	IDUA_MOUSE	ALPHA-L-IDURONIDASE PR	5.10e+01

ALIGNMENTS

RESULT 1  
ID P53\_EQUAS STANDARD; PRT; 207 AA.  
AC Q29480;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).  
GN TP53.  
OS Equus asinus (Donkey).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Perissodactyla; Equidae; Equus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 96342529.  
RA NASIR L., REID S.W.;  
RT "Nucleotide sequence of exons 5 to 9 of the p53 tumour-suppressor  
gene of the donkey (Equus asinus).";  
RL DNA Seq. 6:61-63(1995).  
CC -1- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES  
GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL  
CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN  
TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION, IT IS A  
TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION  
BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF  
THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.  
CC APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF  
BAX AND BCL-2 ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2  
EXPRESSION (BY SIMILARITY).  
CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
CC -1- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY  
OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED  
IN MANY TYPES OF CANCER.  
CC -1- SIMILARITY: BELONGS TO THE P53 FAMILY.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (see http://www.isb-sib.ch/announce/  
or send an email to license@isb-sib.ch).  
-----  
CC ENBL: U26741; AAB41265.1; -  
DR HSP: P04637; 1TSR.  
DR PROSITE; PS00348; P53; 1.  
DR PFAM; PF00870; P53; 1.  
KW Anti-oncogene; DNA-binding; Transcription regulation; Activator;

KW Nuclear protein; Phosphorylation; Apoptosis.  
 FT NON\_TER 1  
 FT DOMAIN 187 199 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 FT NON\_TER 207  
 SQ SEQUENCE 207 AA; 23428 MW; 0FBAE9C1 CRC32;  
 Query Match 100.0%; Score 64; DB 1; Length 207;  
 Best Local Similarity 100.0%; Pred. No. 9.32e-04;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 125 RPILTIITL 133  
 QY 1 RPILTIITL 9  
 RESULT 2  
 ID P53\_HORSE STANDARD; PRT; 280 AA.  
 AC F79892; Q29481;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).  
 GN TP53 OR P53.  
 OS Equus caballus (Horse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Perissodactyla; Equidae; Equus.  
 RN [1]  
 RP SEQUENCE OF 1-263 FROM N.A.  
 RC TISSUE-SPLEEN;  
 RX MEDLINE; 97070350.  
 RA FAZZI K.A., KRAEGL S.A., GRIFFEY S.M., THEON A.P., MADEWELL B.R.;  
 RT "Analysis of the equine tumor suppressor gene p53 in the normal horse  
 and in eight cutaneous squamous cell carcinomas.";  
 RL Cancer Lett. 107:125-130(1996).  
 RN [2]  
 RP SEQUENCE OF 76-280 FROM N.A.  
 RX MEDLINE; 96293865.  
 RA NASIR L., REID S.W.;  
 RT "Nucleotide sequence of exons 5 to 9 of the p53 tumour-suppressor  
 gene of the horse (Equus caballus).";  
 RL DNA Seq. 6:185-187(1996).  
 CC -!- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES  
 CC GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL  
 CC CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN  
 CC TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION, IT IS A  
 CC TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION  
 CC BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF  
 CC THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.  
 CC APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF  
 CC BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2  
 CC EXPRESSION (BY SIMILARITY).  
 CC -!- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -!- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY  
 CC OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED  
 CC IN MANY TYPES OF CANCER.  
 CC -!- SIMILARITY: BELONGS TO THE P53 FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; S83123; AAB46899.1; -;  
 CC EMBL; U37120; AAB18936.1; -;  
 CC HSSP; P04637; 1SAH.  
 CC PROSITE; PS00348; P53; 1.  
 CC PFAM; PF00870; P53; 1.  
 CC Anti-oncogene; DNA-binding; Transcription regulation; Activator;  
 KW Nuclear protein; Phosphorylation; Apoptosis.  
 FT NON\_TER 1

FT DOMAIN 262 274 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 FT CONFLICT 79 79 T -> A (IN REF. 2).  
 FT CONFLICT 83 83 L -> M (IN REF. 2).  
 FT CONFLICT 111 111 A -> V (IN REF. 2).  
 FT CONFLICT 138 138 G -> A (IN REF. 2).  
 FT NON\_TER 280  
 SQ SEQUENCE 280 AA; 30985 MW; B494F872 CRC32;  
 Query Match 100.0%; Score 64; DB 1; Length 280;  
 Best Local Similarity 100.0%; Pred. No. 9.32e-04;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 200 RPILTIITL 208  
 QY 1 RPILTIITL 9  
 RESULT 3  
 ID P53\_SPEBE STANDARD; PRT; 314 AA.  
 AC Q64662;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE CELLULAR TUMOR ANTIGEN P53 (FRAGMENT).  
 GN TP53.  
 OS Sperophilus beecheyi (Beechey ground squirrel).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Rodentia; Sciurognathi; Sciuridae; Scuriinae; Sperophilus.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-THYMUS;  
 RX MEDLINE; 95007566.  
 RA RIVKINA M.B., CULLEN J.M., ROBINSON W.S., MARION P.L.;  
 RT "State of the p53 gene in hepatocellular carcinomas of ground  
 squirrels and woodchucks with past and ongoing infection with  
 RT hepadnaviruses";  
 RL Cancer Res. 54:5430-5437(1994).  
 CC -!- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES  
 CC GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL  
 CC CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN  
 CC TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION, IT IS A  
 CC TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION  
 CC BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF  
 CC THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.  
 CC APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF  
 CC BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2  
 CC EXPRESSION.  
 CC -!- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -!- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY  
 CC OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED  
 CC IN MANY TYPES OF CANCER.  
 CC -!- SIMILARITY: BELONGS TO THE P53 FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; U43902; AAB85628.1; -;  
 CC HSSP; P04637; 1YCS.  
 CC PROSITE; PS00348; P53; 1.  
 CC PFAM; PF00870; P53; 1.  
 CC Anti-oncogene; DNA-binding; Transcription regulation; Activator;  
 KW Nuclear protein; Phosphorylation.  
 FT NON\_TER 1  
 FT DOMAIN 289 301  
 FT CONFLICT 314 314  
 SQ SEQUENCE 314 AA; 34618 MW; D07F433B CRC32;  
 Query Match 100.0%; Score 64; DB 1; Length 314;

Best Local Similarity 100.0%; Pred. No. 9.32e-04; Mismatches 0; Indels 0; Gaps 0;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 227 RPILTIITL 235  
| | | | | | | | | |  
QY 1 RPILTIITL 9

RESULT 4  
ID P53\_XENLA STANDARD; PRT; 363 AA.  
AC P07193;  
DT 01-APR-1988 (Rel. 07, Created)  
DT 01-APR-1988 (Rel. 07, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE CELLULAR TUMOR ANTIGEN P53.  
GN TP53.  
OS Xenopus laevis (African clawed frog).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Amphibia;  
OC Batrachia; Anura; Mesobatrachia; Pipidae; Xenopodinae;  
OC Xenopus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 88143684.  
RA SOUSSI T., DE FROMENTEL C.C., MECHALI M., MAY P., KRESS M.;  
RT "Cloning and characterization of a cDNA from Xenopus laevis coding  
for a protein homologous to human and murine p53.";  
RL Oncogene 1:71-78(1987).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 94134403.  
RA HOEVER M., CLEMENT J.H., WEDLICH D., MONTENARH M., KNOCHEL W.;  
RT "Overexpression of wild-type p53 interferes with normal development  
in Xenopus laevis embryos.";  
RL Oncogene 9:109-120(1994).  
CC -1- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES  
GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL  
CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN  
TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION, IT IS A  
TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION  
BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF  
THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.  
APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF  
BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2  
EXPRESSION (BY SIMILARITY).  
CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
CC -1- TISSUE SPECIFICITY: UBQUITOUS.  
CC -1- SIMILARITY: BELONGS TO THE P53 FAMILY.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL: M36962; AAA49923.1; -  
DR EMBL: X05191; CAA28821.1; -  
DR EMBL: X77546; CAA54672.1; -  
DR EMBL: S68353; AAC60746.1; -  
DR PIR: A29376; A29376.  
DR HSSP: P04637; ITRSR.  
DR PFAM: PF00870; P53; 1.  
KW Anti-oncogene; DNA-binding; Transcription regulation; Activator;  
KW Nuclear protein; Phosphorylation; Apoptosis.  
FT DOMAIN 281 293 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
FT MOD\_RES 362 362 PHOSPHORYLATION (BY SIMILARITY).  
FT CONFLICT 52 52 T -> S (IN REF. 2).  
FT CONFLICT 71 71 MISSING (IN REF. 2).  
FT CONFLICT 296 296 MISSING (IN REF. 2).  
SQ SEQUENCE 363 AA; 40692 MW; 75D7D796 CRC32;

Query Match 100.0%; Score 64; DB 1; Length 363;  
Best Local Similarity 100.0%; Pred. No. 9.32e-04;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 224 RPILTIITL 232  
| | | | | | | | | |  
QY 1 RPILTIITL 9

RESULT 5  
ID P53\_BRARE STANDARD; PRT; 373 AA.  
AC P79734;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE CELLULAR TUMOR ANTIGEN P53.  
GN TP53.  
OS Brachydanio rerio (Zebrafish) (Zebra danio).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Actinopterygii;  
OC Neopterygii; Teleostei; Euteleostei; Ostariophysi; Cypriniformes;  
OC Cyprinidae; Cyprininae; Rasbora; Danio.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 97344388.  
RA CHENG R., FORD B.L., O'NEAL P.E., MATHEWS C.2., BRADFORD C.S.,  
RA THONGTAN T., BARNES D.W., HENDRICKS J.D., BAILEY G.S.;  
RT "Zebrafish (Danio rerio) p53 tumor suppressor gene: cDNA sequence and  
expression during embryogenesis.";  
RL Mol. Mar. Biol. Biotechnol. 6:88-97(1997).  
CC -1- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES  
GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL  
CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN  
TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION, IT IS A  
TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION  
BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF  
THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.  
APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF  
BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2  
EXPRESSION (BY SIMILARITY).  
CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
CC -1- SIMILARITY: BELONGS TO THE P53 FAMILY.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL: U60804; AAB40617.1; -  
DR HSSP: P04637; ITRSR.  
DR ZFIN: ZDB-GENE-990415-32; TP53.  
DR PROSITE: PS00348; P53; 1.  
DR PFAM: PF00870; P53; 1.  
KW Anti-oncogene; DNA-binding; Transcription regulation; Activator;  
KW Nuclear protein; Phosphorylation; Apoptosis.  
FT DOMAIN 280 296 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
FT MOD\_RES 372 372 PHOSPHORYLATION (BY SIMILARITY).  
SQ SEQUENCE 373 AA; 41899 MW; 706A4B9C CRC32;

Query Match 100.0%; Score 64; DB 1; Length 373;  
Best Local Similarity 100.0%; Pred. No. 9.32e-04;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 217 RPILTIITL 225  
| | | | | | | | | |  
QY 1 RPILTIITL 9

RESULT 6  
ID P53\_CANFA STANDARD; PRT; 381 AA.  
AC Q29537;

DT 01-NOV-1997 (Rel. 35, Created)  
DT 15-DEC-1998 (Rel. 37, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE CELLULAR TUMOR ANTIGEN P53.  
GN TP53 OR P53.  
OS Canis familiaris (Dog).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Carnivora; Fissipedia; Canidae; Canis.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=LEUKOCYTE;  
RX MEDLINE; 98178696.  
RA VELDHOFEN N., MILNER J.;  
RT "Isolation of canine p53 cDNA and detailed characterization of the  
RT full length canine p53 protein.";  
RL Oncogene 16:1077-1084(1998).  
RN [2]  
RP SEQUENCE OF 25-300 FROM N.A.  
RC STRAIN=BEAGLE;  
RX MEDLINE; 95323915.  
RA KRAEGLER S.A., PAZZI K.A., MADEWELL B.R.;  
RT "Sequence analysis of canine p53 in the region of exons 3-8.";  
RL Cancer Lett. 92:181-186(1995).  
CC -!- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES  
CC GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL  
CC CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN  
CC TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION, IT IS A  
CC TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION  
CC BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF  
CC THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.  
CC APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF  
CC BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2  
CC EXPRESSION.  
CC -!- SUBCELLULAR LOCATION: NUCLEAR.  
CC -!- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY  
CC OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED  
CC IN MANY TYPES OF CANCER.  
CC -!- SIMILARITY: BELONGS TO THE P53 FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AF060514; AAC16909.1; -.  
DR EMBL; S77819; AAB42022.1; -.  
DR HSSP; P04637; LYCS.  
DR PROSITE; PS00348; P53; 1.  
DR PFAM; PF00870; P53; 1.  
KW Anti-oncogene; DNA-binding; Transcription regulation; Activator;  
KW Nuclear protein; Phosphorylation; Apoptosis.  
FT DOMAIN 1 59 ASP/GLU-RICH (ACIDIC).  
FT DOMAIN 68 137 HYDROPHOBIC.  
FT DOMAIN 307 381  
FT DOMAIN 299 311  
FT MOD\_RES 380 380 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
FT SEQUENCE 381 AA; 42486 MW; 70210863 CRC32;  
Query Match 100.0%; Score 64; DB 1; Length 381;  
Best Local Similarity 100.0%; Pred. No. 9.32e-04;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Db 237 RPILITITL 245  
QY 1 RPILITITL 9  
RESULT 7  
ID P53\_SHEEP STANDARD; PRT; 382 AA.  
PRT;  
SEQUENCE FROM N.A.  
ID P53\_SHEEP STANDARD; PRT; 382 AA.  
PRT;  
SEQUENCE FROM N.A.

P51664;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE CELLULAR TUMOR ANTIGEN P53.  
GN TP53.  
OS Ovis aries (Sheep).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae;  
OC Caprinae; Ovis.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=BLOOD;  
RX MEDLINE; 95352828.  
RA DEQUIEDT F., KETTMANN R., BURNY A., WILLEMS L.;  
RT "Nucleotide sequence of the ovine P53 tumor-suppressor cDNA and its  
RT genomic organization.";  
RL DNA Seq. 5:253-259(1995).  
CC -!- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES  
CC GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL  
CC CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN  
CC TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION, IT IS A  
CC TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION  
CC BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF  
CC THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.  
CC APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF  
CC BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2  
CC EXPRESSION.  
CC -!- SUBCELLULAR LOCATION: NUCLEAR.  
CC -!- SIMILARITY: BELONGS TO THE P53 FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; X81705; CAA57349.1; -.  
DR HSSP; P04637; 1PET.  
DR PROSITE; PS00348; P53; 1.  
DR PFAM; PF00870; P53; 1.  
KW Anti-oncogene; DNA-binding; Transcription regulation; Activator;  
KW Nuclear protein; Phosphorylation; Apoptosis.  
FT DOMAIN 1 66 ASP/GLU-RICH (ACIDIC).  
FT DOMAIN 300 312 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
FT MOD\_RES 381 381 PHOSPHORYLATION (BY SIMILARITY).  
FT SEQUENCE 382 AA; 42809 MW; 0CB99A00 CRC32;  
Query Match 100.0%; Score 64; DB 1; Length 382;  
Best Local Similarity 100.0%; Pred. No. 9.32e-04;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Db 238 RPILITITL 246  
QY 1 RPILITITL 9  
RESULT 8  
ID P53\_BOVIN STANDARD; PRT; 386 AA.  
AC Q29628;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 15-JUL-1998 (Rel. 36, Last annotation update)  
DE CELLULAR TUMOR ANTIGEN P53.  
GN TP53.  
OS Bos taurus (Bovine), and Bos indicus (Zebu).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae;  
OC Bovinae; Bos.  
RN [1]  
RP SEQUENCE FROM N.A.

RC SPECIES=BOVINE; TISSUE=LIVER;  
RX MEDLINE; 95352829.  
RA DEQUIEDT F., KETTMANN R., BURNY A., WILLEMS L.;  
RT "Nucleotide sequence of the bovine p53 tumor-suppressor cDNA.";  
RL DNA Seq. 5:261-264(1995).  
[2]  
RN SEQUENCE OF 13-386 FROM N.A.  
RP SPECIES=BOVINE; STRAIN=HOLSTEIN; TISSUE=THYMUS;  
RX MEDLINE; 96401400.  
RA KOMORI H., ISHIGURO N., HORIUCHI M., SHINAGAWA M., AIDA Y.;  
RT "Predominant p53 mutations in enzootic bovine leukemic cell lines.";  
RL Vet. Immunol. Immunopathol. 52:53-63(1996).  
[3]  
RN SEQUENCE FROM N.A.  
RP SPECIES=B. INDICUS; STRAIN=BORAN; TISSUE=BLOOD;  
RA BISHOP R.R.P., GOBRIGHT E.E.I.;  
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES  
CC GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL  
CC CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN  
CC TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION, IT IS A  
CC TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION  
CC BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF  
CC THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.  
CC APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF  
CC BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2  
CC EXPRESSION.  
CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
CC -1- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY  
CC OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED  
CC IN MANY TYPES OF CANCER.  
CC -1- SIMILARITY: BELONGS TO THE P53 FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; X81704; CAA57348.1; -;  
DR EMBL; D49825; BAA08629.1; -;  
DR EMBL; U74486; AAB51214.1; -;  
DR HSSP; P04637; LYCR.  
DR PROSITE; PS00348; P53; 1.  
DR PFAM; PF00870; P53; 1.  
KW Anti-oncogene; DNA-binding; Transcription regulation; Activator;  
KW Nuclear protein; Phosphorylation; Apoptosis.  
FT DOMAIN 1 59 ASP/GLU-RICH (ACIDIC).  
FT DOMAIN 304 316 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
FT MOD\_RES 385 385 PHOSPHORYLATION (BY SIMILARITY).  
FT CONFLICT 380 380 R -> T (IN REF. 2).  
SQ SEQUENCE 386 AA; 43255 MW; 0322BF3D CRC32;  
Query Match 100.0%; Score 64; DB 1; Length 386;  
Best Local Similarity 100.0%; Pred. No. 9.32e-04;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Db 242 RPILITITL 250  
| | | | | | | | | |  
QY 1 RPILITITL 9  
RESULT 9  
ID P53\_MOUSE STANDARD; PRT; 390 AA.  
AC P02340;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 01-NOV-1990 (Rel. 16, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE CELLULAR TUMOR ANTIGEN P53.  
GN TP53 OR TRP53 OR P53.  
OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Mammalia;  
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
[1]  
RN SEQUENCE FROM N.A.  
RX MEDLINE; 85027173.  
RA BIENZ B., ZAKUT-HOURI R., GIVOL D., OREN M.;  
RT "Analysis of the gene coding for the murine cellular tumour antigen  
RT p53.";  
RL EMBO J. 3:2179-2183(1984).  
[2]  
RN SEQUENCE FROM N.A.  
RX MEDLINE; 84068204.  
RA ZAKUT-HOURI R., OREN M., BIENZ B., LAVIE V., HAZUM S., GIVOL D.;  
RT "A single gene and a pseudogene for the cellular tumour antigen p53.";  
RL Nature 306:594-597(1983).  
[3]  
RN SEQUENCE FROM N.A.  
RX MEDLINE; 84272240.  
RA JENKINS J.R., RUDGE K., REDMOND S., WADE-EVANS A.;  
RT "Cloning and expression analysis of full length mouse cDNA sequences  
RT encoding the transformation associated protein p53.";  
RL Nucleic Acids Res. 12:5609-5626(1984).  
[4]  
RN SEQUENCE FROM N.A. (CLONES PCD53; P53-M11 AND P53-M8).  
RX MEDLINE; 87064640.  
RA ARAI N., NOMURA D., YOKOTA K., WOLF D., BRILL E., SHOHAT O.;  
RT "Immunologically distinct p53 molecules generated by alternative  
RT splicing.";  
RL Mol. Cell. Biol. 6:3232-3239(1986).  
[5]  
RN SEQUENCE OF 222-258 FROM N.A.  
RX MEDLINE; 92115342.  
RA BURNS P.A., KEMP C.J., GANNON J.V., LANE D.P., BRENNER R.,  
RA BALMAIN A.;  
RT "Loss of heterozygosity and mutational alterations of the p53 gene in  
RT skin tumours of interspecific hybrid mice.";  
RL Oncogene 6:2363-2369(1991).  
[6]  
RN PHOSPHORYLATION SITES.  
RX MEDLINE; 86149247.  
RA SAMAD A., ANDERSON C.W., CARROLL R.B.;  
RT "Mapping of phosphonoester and apparent phosphodiester bonds of the  
RT oncogene product p53 from simian virus 40-transformed 3T3 cells.";  
RL Proc. Natl. Acad. Sci. U.S.A. 83:897-901(1986).  
[7]  
RN PHOSPHORYLATION SITES.  
RX MEDLINE; 91006019.  
RA MEER D.W., SIMON S., KIKAWA U., ECKHART W.;  
RT "The p53 tumour suppressor protein is phosphorylated at serine 389 by  
RT casein kinase II.";  
RL EMBO J. 9:3253-3260(1990).  
CC -1- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES  
CC GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL  
CC CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN  
CC TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION, IT IS A  
CC TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION  
CC BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF  
CC THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.  
CC APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF  
CC BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2  
CC EXPRESSION.  
CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
CC -1- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY  
CC OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED  
CC IN MANY TYPES OF CANCER.  
CC -1- SIMILARITY: BELONGS TO THE P53 FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial





```
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE CELLULAR TUMOR ANTIGEN P53.
GN TP53.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Lagomorpha; Leporidae; Oryctolagus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-NEW ZEALAND;
RX MEDLINE; 97208869.
RA LE GOAS F., MAY P., RONCO P., CARON DE FROMENTEL C.;
RT "CDNA cloning and immunological characterization of rabbit p53.";
RL Gene 185:169-173(1997).
CC -!- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES
CC GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL
CC CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN
CC TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION. IT IS A
CC TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION
CC BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF
CC THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.
CC APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF
CC BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2
CC EXPRESSION (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY
CC OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED
CC IN MANY TYPES OF CANCER.
CC -!- SIMILARITY: BELONGS TO THE P53 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X90592; CAA62216.1; -.
CC HSSP; P04637; 1YCR.
CC PROSITE; PS00348; P53; 1.
CC PFAM; PF00870; P53; 1.
CC KW Anti-oncogene; DNA-binding; Transcription regulation; Activator;
CC Nuclear protein; Phosphorylation; Apoptosis.
CC FT DOMAIN 1 70 ASP/GLU-RICH (ACIDIC).
CC FT DOMAIN 308 321 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
CC FT MOD_RES 390 390 PHOSPHORYLATION (BY SIMILARITY).
CC SQ SEQUENCE 391 AA; 43435 MW; 30A36172 CRC32;
CC -----
CC Query Match 100.0%; Score 64; DB 1; Length 391;
CC Best Local Similarity 100.0%; Pred. No. 9.32e-04;
CC Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CC
CC Db 246 RPILTIITL 254
CC | | | | | | | |
CC QY 1 RPILTIITL 9
CC -----
CC RESULT 12
CC ID P53_MACEFA STANDARD; PRT; 393 AA.
CC AC P56423;
CC DT 15-JUL-1998 (Rel. 36, Created)
CC DT 15-JUL-1998 (Rel. 36, Last sequence update)
CC DT 15-JUL-1998 (Rel. 36, Last annotation update)
CC DE CELLULAR TUMOR ANTIGEN P53.
CC GN TP53 OR P53.
CC OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
CC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
CC OC Eutheria; Primates; Catarrhini; Cercopithecoidea; Cercopithecinae;
CC OC Macaca.
CC RN [1]
CC RP SEQUENCE FROM N.A.
CC RA KHAN M.A., HANSEN C., WELSH J.A., BENNETT W.P.;
CC RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
CC -----
CC -!- FUNCTION: ACT AS A TUMOR SUPPRESSOR IN MANY TUMOR TYPES. INDUCES
CC GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL
CC CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN
CC TUMOR SUPPRESSION. IT ACTS IN CELL CYCLE REGULATION. IT IS A
CC TRANS-ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELLULAR DIVISION
CC BY CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF
CC THE GENES ACTIVATED IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.
CC APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF
CC BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2
CC EXPRESSION (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY
CC OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED
CC IN MANY TYPES OF CANCER.
CC -!- SIMILARITY: BELONGS TO THE P53 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U48957; AAB91535.1; -.
CC HSSP; P04637; ISAH.
CC PROSITE; PS00348; P53; 1.
CC PFAM; PF00870; P53; 1.
CC KW Anti-oncogene; DNA-binding; Transcription regulation; Activator;
CC Nuclear protein; Phosphorylation; Apoptosis.
CC FT DOMAIN 1 80 ASP/GLU-RICH (ACIDIC).
CC FT DOMAIN 81 150 HYDROPHOBIC.
CC FT DOMAIN 319 393 HIGHLY BASIC AND MAY BE INVOLVED IN
CC FT DOMAIN 311 323 NUCLEAR LOCALIZATION SIGNAL.
CC FT MOD_RES 392 392 PHOSPHORYLATION (BY SIMILARITY).
CC SQ SEQUENCE 393 AA; 43678 MW; 2499AC47 CRC32;
CC -----
CC Query Match 100.0%; Score 64; DB 1; Length 393;
CC Best Local Similarity 100.0%; Pred. No. 9.32e-04;
CC Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CC
CC Db 249 RPILTIITL 257
CC | | | | | | | |
CC QY 1 RPILTIITL 9
CC -----
CC RESULT 13
CC ID P53_HUMAN STANDARD; PRT; 393 AA.
CC AC P04637;
CC DT 13-AUG-1987 (Rel. 05, Created)
CC DT 01-MAR-1989 (Rel. 10, Last sequence update)
CC DT 15-JUL-1999 (Rel. 38, Last annotation update)
CC DE CELLULAR TUMOR ANTIGEN P53 (PHOSPHOPROTEIN P53).
CC GN TP53.
CC OS Homo sapiens (Human).
CC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
CC OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
CC RN [1]
CC RP SEQUENCE FROM N.A.
CC RX MEDLINE; 85230577.
CC RA ZAKUT-HOURI R., BIENZ-TADMOR B., GIVOL D., OREN M.;
CC RT "Human p53 cellular tumor antigen: cDNA sequence and expression in
CC COS cells.";
CC RL EMBO J. 4:1251-1255(1985).
CC RN [2]
CC RP SEQUENCE FROM N.A.
CC RX MEDLINE; 87064416.
CC RA LAMB P., CRAWFORD L.;
CC RT "Characterization of the human p53 gene.";
CC RL Mol. Cell. Biol. 6:1379-1385(1986).
CC RN [3]
CC RP SEQUENCE FROM N.A.
```

RX MEDLINE; 85267676.  
RA HARLOW E., WILLIAMSON N.M., RALSTON R., HELFMAN D.M., ADAMS T.E.;  
RT "Molecular cloning and in vitro expression of a cDNA clone for human  
RT cellular tumor antigen p53";  
RL Mol. Cell. Biol. 5:1601-1610(1985).  
RN [4]  
RP TRANSFORMED HYBRIDOMA SV-80 CELL LINE, SEQUENCE FROM N.A.  
RX MEDLINE; 87089826.  
RA HARRIS N., BRILL E., SHOHAT O., PROKOCIMER M., WOLF D., ARAI N.,  
RA ROTTER V.;  
RT "Molecular basis for heterogeneity of the human p53 protein.";  
RL Mol. Cell. Biol. 6:4650-4656(1986).  
RN [5]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 89108008.  
RA BUCHMAN V.L., CHUMAKOV P.M., NINKINA N.N., SAMARINA O.P.;  
RA GEORGIEV G.P.;  
RT "A variation in the structure of the protein-coding region of the  
RT human p53 gene";  
RL Gene 70:245-252(1988).  
RN [6]  
RP SEQUENCE OF 101-393 FROM N.A.  
RX MEDLINE; 85126934.  
RA MATIASHEWSKI G., LAMB P., PIM D., PEACOCK J., CRAWFORD L.,  
RA BENCHIMOL S.;  
RT "Isolation and characterization of a human p53 cDNA clone: expression  
RT of the human p53 gene";  
RL EMBO J. 3:3257-3262(1984).  
RN [7]  
RP NUCLEAR LOCALIZATION SIGNAL.  
RX MEDLINE; 90191730.  
RA ADDISON C., JENKINS J.R., STURZBECHER H.-W.;  
RT "The p53 nuclear localisation signal is structurally linked to a  
RT p34cdc2 kinase motif";  
RL Oncogene 5:423-426(1990).  
RN [8]  
RP PHOSPHORYLATION BY P60/CDC2 AND CYCLIN B/CDC2.  
RX MEDLINE; 90280456.  
RA SCHEIDTMANN K.H., MUMBY M.C., RUNDELL K., WALTER G.;  
RT "Dephosphorylation of simian virus 40 large-T antigen and p53 protein  
RT by protein phosphatase 2A: inhibition by small-t antigen.";  
RL Mol. Cell. Biol. 11:1996-2003(1991).  
RN [10]  
RP STRUCTURE BY NMR OF 319-360.  
RX MEDLINE; 94294808.  
RA CLORE G.M., OMICHINSKI J.G., SAKAGUCHI K., ZAMBRANO N., SAKAMOTO H.,  
RA APPELLA E., GRONENBORN A.M.;  
RT "High-resolution structure of the oligomerization domain of p53 by  
RT multidimensional NMR";  
RL Science 265:386-391(1994).  
RN [11]  
RP STRUCTURE BY NMR OF 325-355.  
RX MEDLINE; 95292092.  
RA LEE W., HARVEY T.S., YIN Y., YAU P., LITCHFIELD D., ARROWSMITH C.H.;  
RT "Solution structure of the tetrameric minimum transforming domain of  
RT p53";  
RL Nat. Struct. Biol. 1:877-890(1994).  
RN [12]  
RP STRUCTURE BY NMR OF 326-354.  
RX MEDLINE; 98026899.  
RA MCCOY M., STAVRIDIS E.S., WATERMAN J.L., WIECZOREK A.M., OPELLA S.J.,  
RA HALAZONETIS T.D.;  
RT "Hydrophobic side-chain size is a determinant of the  
RT three-dimensional structure of the p53 oligomerization domain.";  
RL EMBO J. 16:6230-6236(1997).  
RN [13]  
RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 94-289.

RX MEDLINE; 94294806.  
RA CHO Y., GORINA S., JEFFREY P.D., PAVLETICH N.P.;  
RT "Crystal structure of a p53 tumor suppressor-DNA complex:  
RT understanding tumorigenic mutations";  
RL Science 265:346-355(1994).  
RN [14]  
RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS) OF 13-29 IN COMPLEX WITH MDM2.  
RX MEDLINE; 97081050.  
RA KUSSIE P.H., GORINA S., MARECHAL V., ELEBAAS B., MOREAU J.,  
RA LEVINE A.J., PAVLETICH N.P.;  
RT "Structure of the MDM2 oncoprotein bound to the p53 tumor suppressor  
RT transactivation domain";  
RL Science 274:948-953(1996).  
RN [15]  
RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 97-287 IN COMPLEX WITH 53BP2.  
RX MEDLINE; 97035414.  
RA GORINA S., PAVLETICH N.P.;  
RT "Structure of the p53 tumor suppressor bound to the ankyrin and SH3  
RT domains of 53BP2";  
RL Science 274:1001-1005(1996).  
RN [16]  
RP REVIEW.  
RX MEDLINE; 94090335.  
RA HARRIS C.C.;  
RT "p53: at the crossroads of molecular carcinogenesis and risk  
RT assessment";  
RL Science 262:1980-1981(1993).  
RN [17]  
RP REVIEW ON VARIANTS.  
RX MEDLINE; 91289156.  
RA HOOLSTEIN M., SIDRANSKY D., VOGELSTEIN B., HARRIS C.C.;  
RT "p53 mutations in human cancers";  
RL Science 253:49-53(1991).  
RN [18]  
RP REVIEW ON VARIANTS.  
RX MEDLINE; 96271983.  
RA DE VRIES E.M.G., RICKE D.O., DE VRIES T.N., HARTMANN A., BLASZYK H.,  
RA LIAO D., SOUSIE T., KOVACH J.S., SOMMER S.S.;  
RT "Database of mutations in the p53 and APC tumor suppressor genes  
RT designed to facilitate molecular epidemiological analyses";  
RL Hum. Mutat. 7:202-213(1996).  
RN [19]  
RP VARIANT ARG-72.  
RX MEDLINE; 91153807.  
RA OLSCHWANG S., LAURENT-PUIG P., VASSAL A., SALMON R.-J., THOMAS G.;  
RT "Characterization of a frequent polymorphism in the coding sequence  
RT of the Tp53 gene in colonic cancer patients and a control  
RT population";  
RL Hum. Genet. 86:369-370(1991).  
RN [20]  
RP VARIANT LFS THR-133.  
RX MEDLINE; 92034774.  
RA LAW J.C., STRONG L.C., CHIDAMBARAM A., FERRELL R.E.;  
RT "A germ-line mutation in exon 5 of the p53 gene in an extended cancer  
RT family";  
RL Cancer Res. 51:6385-6387(1991).  
RN [21]  
RP VARIANTS LFS CYS-245; TRP-248; PRO-252 AND LYS-258.  
RX MEDLINE; 91057657.  
RA MALKIN D., LI F.P., STRONG L.C., FRAUMENI J.F. JR., NELSON C.E.,  
RA KIM D.H., KASSEL J., GRIKA M.A., BISCHOFF F.Z., TAINSKY M.A.,  
RA FRIEND S.H.;  
RT "Germ-line p53 mutations in a familial syndrome of breast cancer,  
RT sarcomas, and other neoplasms";  
RL Science 250:1233-1238(1990).  
RN [22]  
RP VARIANT LFS ASP-245.  
RX MEDLINE; 91080929.  
RA SRIVASTAVA S., ZOU Z., PIROLLO K., BLATTNER W., CHANG E.H.;  
RT "Germ-line transmission of a mutated p53 gene in a cancer-prone  
RT family with Li-Fraumeni syndrome";  
RL Nature 348:747-749(1990).  
RN [23]

```
##molecule_type DNA
##residues 1-885 ##label BOH
##cross-references EMBL:274966; NID:g1420196; PID:e252338; PID:g1420197;
MIPS:YOR058c
##experimental_source strain S288c

GENETICS
#gene SGD:ASE1
##cross-references SGD:S000584; MIPS:YOR058c
#map_position 15R
SUMMARY #length 885 #molecular-weight 101623 #checksum 8781

Query Match 76.9%; Score 50; DB 2; Length 885;
Best Local Similarity 44.4%; Pred. No. 5.03e+00;
Matches 4; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 823 EPEHSYKL 831
:|: :|||
Qy 1 QPDDAVYKL 9

RESULT 7
ENTRY #type complete
TITLE mannosyltransferase A (ntfa) homolog - Archaeoglobus fulgidus
ORGANISM #formal_name Archaeoglobus fulgidus
DATE 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change
05-Jun-1998
ACCESSIONS E69255
REFERENCE E69250
#authors Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson, R.J.; Gwinn, M.; Hickey, E.K.; Peterson, J.D.; Richardson, D.L.; Kierlavage, A.R.; Graham, D.E.; Kyrpides, N.C.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.; Dougherty, B.A.; McKenny, K.; Adams, M.D.; Loftus, B.; Peterson, S.; Reich, C.I.; McNeil, L.K.; Badger, J.H.; Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.; Utterback, T.; Cotton, M.D.; Spriggs, T.; Artach, P.; Kaine, B.P.; Sykes, S.M.; Sadow, P.W.; D'Andrea, K.P.; Bowman, C.; Fujii, C.; Garland, S.A.; Mason, T.M.; Olsen, G.J.; Fraser, C.M.; Smith, H.O.; Woese, C.R.; Venter, J.C.
#journal Nature (1997) 390:364-370
#title The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeon Archaeoglobus fulgidus.
#cross-references MUID:98049343
#accession E69255
#status preliminary; nucleic acid sequence not shown; translation not shown
##molecule_type DNA
##residues 1-1213 ##label KLE
##cross-references GB:AE001103; GB:AE000782; NID:g2689426; PID:g2650604; TIGR:AF0045
SUMMARY #length 1213 #molecular-weight 140592 #checksum 4026

Query Match 76.9%; Score 50; DB 2; Length 1213;
Best Local Similarity 62.5%; Pred. No. 5.03e+00;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 730 PNEVYKL 737
:|: :|||
Qy 2 PDDAVYKL 9

RESULT 8
ENTRY #type complete
TITLE secretory component precursor - rabbit
ALTERNATE_NAMES poly-ig receptor; polymorphic immunoglobulin receptor
CONTAINS free secretory component; transmembrane secretory component
ORGANISM #formal_name Oryctolagus cuniculus #common_name domestic rabbit
DATE 15-Nov-1984 #sequence_revision 15-Nov-1984 #text_change
25-Oct-1996
ACCESSIONS A02111; A28077
```

```
REFERENCE A02111
#authors Mostov, K.E.; Friedlander, M.; Blobel, G.
#journal Nature (1984) 308:37-43
#title The receptor for transepithelial transport of IgA and IgM contains multiple immunoglobulin-like domains.
#cross-references MUID:84142246
#accession A02111
##molecule_type mRNA
##residues 1-773 ##label MOS
##note the authors translated the codon ACC for residue 54 as Asn

REFERENCE A28077
#authors Frutiger, S.; Hughes, G.J.; Hanly, W.C.; Jaton, J.C.
#journal J. Biol. Chem. (1988) 263:8120-8125
#title Rabbit secretory components of different allotypes vary in their carbohydrate content and their sites of N-linked glycosylation.
#cross-references MUID:88228032
#accession A28077
##molecule_type protein
##residues 87-114; 410-424 ##label FRU
COMMENT This receptor binds polymeric IgA and IgM at the basolateral surface of epithelial cells. The complex is then transported across the cell to be secreted at the apical surface. During this process, cleavage occurs to separate the extracellular portion, also known as the secretory component, from the transmembrane segment.
COMMENT The five domains exhibit homology with immunoglobulin V regions. The similarity is strongest between the fourth domain and kappa chain V regions.
COMMENT Alternative splicing in the extracellular domain leads to high or low molecular weight forms of secretory component.
CLASSIFICATION #superfamily secretory component; immunoglobulin homology
KEYWORDS alternative splicing; duplication; glycoprotein; immunoglobulin receptor; polymorphism; transcytosis; transmembrane protein

FEATURE
1-18 #domain signal sequence #status predicted #label SIG\
19-773 #product transmembrane secretory component #status predicted #label MATM\
19-575 #product free secretory component #status predicted #label MATF\
30-647 #domain extracellular #status predicted #label EXT\
39-117 #domain immunoglobulin homology #label IG1\
148-227 #domain immunoglobulin homology #label IG2\
253-326 #domain immunoglobulin homology #label IG3\
362-440 #domain immunoglobulin homology #label IG4\
471-540 #domain immunoglobulin homology #label IG5\
648-670 #domain immunoglobulin homology #label TMM\
671-773 #domain intracellular #status predicted #label INT\
46-115,155-225, #disulfide_bonds #status predicted\
260-324,369-438, #binding_site carbohydrate (Asn) (covalent) (partial)\
478-538 #status experimental\
108 #binding_site carbohydrate (Asn) (covalent) #status experimental
418 #length 773 #molecular-weight 83886 #checksum 7723

SUMMARY
Query Match 75.4%; Score 49; DB 1; Length 773;
Best Local Similarity 75.0%; Pred. No. 8.13e+00;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 137 EPDDVYKL 144
:|: :|||
Qy 1 QPDDAVYKL 8

RESULT 9
ENTRY #type complete
TITLE hypothetical protein - Synecocystis sp. (strain PCC 6803)
ORGANISM #formal_name Synecocystis sp.
#variety PCC 6803
```

DATE 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change  
21-Aug-1998  
ACCESSIONS S76815  
REFERENCE S74322  
#authors Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; Hirose, M.; Sugita, M.; Sasamoto, S.; Kimura, T.; Hosouchi, T.; Matsuno, A.; Muraki, A.; Nakazaki, N.; Naruo, K.; Okumura, S.; Shimpō, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda, M.; Tabata, S.  
#journal DNA Res. (1996) 3:109-136  
#title Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions.  
#cross-references MUID:97061201  
#accession S76815  
##status preliminary  
##molecule\_type DNA  
##residues 1-832 #label KAN  
##cross-references EMBL:D90916; GB:AB001339; NID:g1653715; PID:d1019460; PID:g1653816  
##note the nucleotide sequence was submitted to the EMBL Data Library, June 1996  
SUMMARY #length 832 #molecular-weight 92864 #checksum 8113  
Query Match 75.4%; Score 49; DB 2; Length 832;  
Best Local Similarity 44.4%; Pred. No. 8.13e+00;  
Matches 4; Conservative 4; Mismatches 1; Indels 0; Gaps 0;  
Db 309 QPDEQIFRL 317  
|||: |||  
QY 1 QPDDAVYKL 9  
RESULT 10  
ENTRY S64506 #type complete  
TITLE protein kinase BUB1 (EC 2.7.1.1), checkpoint-associated - yeast (Saccharomyces cerevisiae)  
ALTERNATE\_NAMES protein G7542; protein YGR188c  
ORGANISM #formal\_name Saccharomyces cerevisiae  
DATE 17-May-1996 #sequence\_revision 17-May-1996 #text\_change  
ACCESSIONS S64506; A56354; S50224  
REFERENCE S64499  
#authors Arroyo, J.; Garcia-Gonzalez, M.; Garcia-Saez, M.I.; Sanchez-Perez, M.; Nombela, C.  
#submission submitted to the Protein Sequence Database, May 1996  
#accession S64506  
##molecule\_type DNA  
##residues 1-1021 #label ARR  
##cross-references EMBL:Z72973; NID:g1323333; PID:e243726; PID:g1323334; MIPS:YGR188c  
##experimental\_source strain S288c  
REFERENCE A56354  
#authors Roberts, B.T.; Fart, K.A.; Hoyt, M.A.  
#journal Mol. Cell. Biol. (1994) 14:8282-8291  
#title The *Saccharomyces cerevisiae* checkpoint gene BUB1 encodes a novel protein kinase.  
#cross-references MUID:95059057  
#accession A56354  
##status preliminary  
##molecule\_type DNA  
##residues 1-530, 'v' 532-1021 #label ROB  
##cross-references GB:L32027; NID:g475127; PID:g475128  
GENETICS  
#gene SGD:BUB1  
##cross-references SGD:S0003420; MIPS:YGR188c  
#map\_position 7R  
KEYWORDS autophosphorylation; cell division control; phosphoprotein; phosphotransferase; protein kinase  
SUMMARY #length 1021 #molecular-weight 117867 #checksum 642

Query Match 75.4%; Score 49; DB 2; Length 1021;  
Best Local Similarity 55.6%; Pred. No. 8.13e+00;  
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;  
Db 300 QSNPNPVYKL 308  
|||: |||  
QY 1 QPDDAVYKL 9  
RESULT 11  
ENTRY A54146 #type complete  
TITLE invasion-inducing protein Tiam-1 - mouse  
ORGANISM #formal\_name Mus musculus #common\_name house mouse  
DATE 02-Aug-1994 #sequence\_revision 02-Aug-1994 #text\_change 29-May-1998  
ACCESSIONS A54146  
REFERENCE A54146  
#authors Habets, G.G.M.; Scholtes, E.H.M.; Zuydgeest, D.; van der Kammen, R.A.; Stam, J.C.; Berns, A.; Collard, J.G.  
#journal Cell (1994) 77:537-549  
#title Identification of an invasion-inducing gene, Tiam-1, that encodes a protein with homology to GDP-GTP exchangers for Rho-like proteins.  
#accession A54146  
##status preliminary  
##molecule\_type mRNA  
##residues 1-1591 #label HAB  
##cross-references GB:U05245; NID:g497638; PID:g497639  
CLASSIFICATION #superfamily CDC24 homology; pleckstrin repeat homology  
FEATURE 1040-1234 #domain CDC24 homology #label CD24  
SUMMARY #length 1591 #molecular-weight 177532 #checksum 3127  
Query Match 75.4%; Score 49; DB 2; Length 1591;  
Best Local Similarity 44.4%; Pred. No. 8.13e+00;  
Matches 4; Conservative 3; Mismatches 2; Indels 0; Gaps 0;  
Db 823 QPEDDIYEL 831  
|||: |||  
QY 1 QPDDAVYKL 9  
RESULT 12  
ENTRY A33378 #type complete  
TITLE fasciclin III precursor - fruit fly (*Drosophila melanogaster*)  
ORGANISM #formal\_name *Drosophila melanogaster*  
DATE 21-Feb-1990 #sequence\_revision 21-Feb-1990 #text\_change 24-Sep-1998  
ACCESSIONS A33378  
REFERENCE A33378  
#authors Snow, P.M.; Bieber, A.J.; Goodman, C.S.  
#journal Cell (1989) 59:313-323  
#title Fasciclin III: a novel homophilic adhesion molecule in *Drosophila*.  
#cross-references MUID:90030406  
#accession A33378  
##status preliminary  
##molecule\_type mRNA  
##residues 1-508 #label SNO  
##cross-references GB:M27813; NID:g157423; PID:g157424  
GENETICS  
#gene FlyBase:Fas3  
##cross-references FlyBase:FBgn0000636  
KEYWORDS phosphoprotein; transmembrane protein  
SUMMARY #length 508 #molecular-weight 55883 #checksum 7642  
Query Match 73.8%; Score 48; DB 2; Length 508;  
Best Local Similarity 77.8%; Pred. No. 1.30e+01;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
Db 240 QPDAAVYGL 248  
|||: |||  
QY 1 QPDDAVYKL 9

```
RESULT 13
ENTRY
TITLE      S27387      #type complete
ORGANISM   Interferon alpha receptor type 1 precursor - bovine
DATE       #formal_name Bos primigenius taurus #common_name cattle
           13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change
           13-Nov-1998
ACCESSIONS S27387; S33770
REFERENCE   S27387
#authors   Mouchel-Vielh, E.; Lutfalla, G.; Mogensen, K.E.; Uze, G.
#journal   FEBS Lett. (1992) 313:255-259
#title     Specific antiviral activities of the human alpha interferons
           are determined at the level of receptor (IFNAR) structure.
#cross-references MUID:93076908
#accession S27387
##status   preliminary; nucleic acid sequence not shown
##molecule_type mRNA
##residues 1-560 #label MOU
##cross-references EMBL:X68443; NID:g431; PID:g432
##experimental_source MDBK cells
REFERENCE   S33770
#authors   Lim, J.K.; Langer, J.A.
#journal   Blochim. Biophys. Acta (1993) 1173:314-319
#title     Cloning and characterization of a bovine alpha interferon
           receptor.
#cross-references MUID:93305725
#accession S33770
##status   preliminary; nucleic acid sequence not shown
##molecule_type mRNA
##residues 1-421, 'V' 423-560 #label LIM
##cross-references EMBL:L06320; NID:g163187; PID:g163188
##experimental_source lung
KEYWORDS   antiviral; cytokine receptor; transmembrane protein
FEATURE
1-24
25-560     #domain signal sequence #status predicted #label SIG\
           #product interferon alpha receptor type 1 #status
           predicted #label MAT
SUMMARY    #length 560 #molecular-weight 63818 #checksum 4991
Query Match 73.8%; Score 48; DB 2; Length 560;
Best Local Similarity 62.5%; Pred. No. 1.30e+01;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 185 PEDKIYKL 192
QY 2 PDDAVYKL 9

RESULT 14
ENTRY
TITLE      A27450      #type complete
ORGANISM   Olfactory marker protein - rat
DATE       #formal_name Rattus norvegicus #common_name Norway rat
           31-Dec-1988 #sequence_revision 16-Feb-1996 #text_change
           20-Mar-1998
ACCESSIONS A27450; A55025
REFERENCE   A27450
#authors   Rogers, K.E.; Dasgupta, P.; Gubler, U.; Grillo, M.;
           Khew-Goodall, Y.S.; Margolis, F.L.
#journal   Proc. Natl. Acad. Sci. U.S.A. (1987) 84:1704-1708
#title     Molecular cloning and sequencing of a cDNA for olfactory
           marker protein.
#cross-references MUID:87175546
#accession A27450
##molecule_type mRNA
##residues 1-162 #label ROG
##cross-references GB:M15644; NID:g205849; PID:g205850
REFERENCE   A55025
#authors   Sydor, W.; Teitelbaum, Z.; Blacher, R.; Sun, S.; Benz, W.;
           Margolis, F.L.
#journal   Arch. Biochem. Biophys. (1986) 249:351-362
#title     Amino acid sequence of a unique neuronal protein: rat
           olfactory marker protein.
```

```
#accession A55025 preliminary
##status preliminary
##molecule_type protein
##residues 1-162 #label SYD
KEYWORDS   acetylated amino end
FEATURE
1          #modified_site acetylated amino end (Ala) #status
           experimental
SUMMARY    #length 162 #molecular-weight 18721 #checksum 1471
Query Match 72.3%; Score 47; DB 2; Length 162;
Best Local Similarity 44.4%; Pred. No. 2.07e+01;
Matches 4; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 46 RPAESVYRL 54
QY 1 QPDDAVYKL 9

RESULT 15
ENTRY
TITLE      B54261      #type complete
ORGANISM   Olfactory marker protein - mouse
DATE       #formal_name Mus musculus #common_name house mouse
           09-Sep-1994 #sequence_revision 09-Sep-1994 #text_change
           10-Sep-1997
ACCESSIONS B54261; I48878
REFERENCE   B54261
#authors   Buiakova, O.I.; Krishna, N.S.R.; Getchell, T.V.; Margolis,
           F.L.
#journal   Genomics (1994) 20:452-462
#title     Human and rodent OMP genes: conservation of structural and
           regulatory motifs and cellular localization.
#accession B54261
##status preliminary
##molecule_type DNA
##residues 1-163 #label BUI
##cross-references GB:U01213; NID:g457940; PID:g520741
REFERENCE   I48878
#authors   Brown, K.A.; Sutcliffe, M.J.; Steele, K.; Brown, S.D.
#journal   Mamm. Genome (1994) 5:11-14
#title     Sequencing of the Olfactory Marker Protein Gene in Normal and
           Shaker-1 Mutant Mice.
#cross-references MUID:94154378
#accession I48878
##status preliminary; translated from GB/EMBL/DBJ
##molecule_type DNA
##residues 1-163 #label RES
##cross-references EMBL:U02557; NID:g493516; PID:g493517
SUMMARY    #length 163 #molecular-weight 18866 #checksum 5346
Query Match 72.3%; Score 47; DB 2; Length 163;
Best Local Similarity 44.4%; Pred. No. 2.07e+01;
Matches 4; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 47 RPAESVYRL 55
QY 1 QPDDAVYKL 9

Search completed: Fri Apr 14 23:37:23 2000
Job time : 11 secs.
```

THIS PAGE BLANK (USPTO)

\*\*\*\*\*  
M P S R L H  
\*\*\*\*\*  
(TM)  
\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Apr 14 23:37:41 2000; Maspar time 5.70 Seconds  
Tabular output not generated.  
47.163 Million cell updates/sec

Title: >US-08-452-843-7  
Description: (1-9) from US08452843.pap  
Perfect Score: 65  
Sequence: 1 QPDDAVYKL 9

Scoring table: PAM 150  
Gap 15

Searched: 82229 seqs, 2986486 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: swiss-prot38  
1:swissprot

Statistics: Mean 23.823; Variance 24.828; scale 0.960

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	53	81.5	162	1	OMP_HUMAN OLFACATORY MARKER PROTE	3.17e+01
2	50	76.9	885	1	ASEL_YEAST ANAPHASE SPINDLE ELONG	1.65e+00
3	49	75.4	262	1	RS4_CANAL 40S RIBOSOMAL PROTEIN	2.80e+00
4	49	75.4	773	1	PIGR_RABIT POLYMERIC-IMMUNOGLOBUL	2.80e+00
5	49	75.4	1021	1	BUBL_YEAST CHECKPOINT SERINE/THRE	2.80e+00
6	49	75.4	1591	1	TIAM_MOUSE T-LYMPHOMA INVASION AN	2.80e+00
7	49	75.4	1591	1	TIAM_MOUSE T-LYMPHOMA INVASION AN	2.80e+00
8	48	73.8	218	1	Y4VH_RHISN HYPOTHETICAL 24.6 KD P	4.74e+00
9	48	73.8	508	1	PAS3_DRONE FASCICLIN III PRECURSOR	4.74e+00
10	48	73.8	560	1	INRL_SHEEP INTERFERON-ALPHA/BETA	4.74e+00
11	48	73.8	560	1	INRL_BOVIN INTERFERON-ALPHA/BETA	4.74e+00
12	47	72.3	162	1	OMP_RAT OLFACATORY MARKER PROTE	7.94e+00
13	47	72.3	447	1	G6PD_SCHPO GLUCOSE-6-PHOSPHATE 1-	7.94e+00
14	47	72.3	522	1	G6PD_CAEEL GLUCOSE-6-PHOSPHATE 1-	7.94e+00
15	47	72.3	563	1	VVFH_BACSU PUTATIVE L-LACTATE PER	7.94e+00
16	47	72.3	574	1	G6PC_SPIOL GLUCOSE-6-PHOSPHATE 1-	7.94e+00
17	47	72.3	577	1	G6PC_SOLTU GLUCOSE-6-PHOSPHATE 1-	7.94e+00
18	47	72.3	2211	1	FA5_BOVIN COAGULATION FACTOR V P	7.94e+00
19	46	70.8	308	1	P2A3_YEAST SERINE/THREONINE PROTE	1.32e+01
20	46	70.8	362	1	OGRE_DRONE OGRE LOCUS PROTEIN.	1.32e+01
21	46	70.8	434	1	UROK_CHICK UROKINASE-TYPE PLASMIN	1.32e+01
22	46	70.8	512	1	G6P2_MOUSE GLUCOSE-6-PHOSPHATE 1-	1.32e+01
23	46	70.8	514	1	G6PD_RAT GLUCOSE-6-PHOSPHATE 1-	1.32e+01

24	46	70.8	514	1	G6P1_MOUSE GLUCOSE-6-PHOSPHATE 1-	1.32e+01
25	46	70.8	514	1	G6PD_HUMAN GLUCOSE-6-PHOSPHATE 1-	1.32e+01
26	46	70.8	514	1	G6PD_MACRO GLUCOSE-6-PHOSPHATE 1-	1.32e+01
27	46	70.8	550	1	THDH_ARXAD THREONINE DEHYDRATASE	1.32e+01
28	46	70.8	656	1	VEXE_SALT1 VI POLYSACCHARIDE EXPO	1.32e+01
29	46	70.8	697	1	TGLC_CHICK PROTEIN-GLUTAMINE GAMM	1.32e+01
30	45	69.2	171	1	BAR_STRCO PHOSPHINOTHRICIN ACETY	2.17e+01
31	45	69.2	194	1	RS4_BOVIN 40S RIBOSOMAL PROTEIN	2.17e+01
32	45	69.2	262	1	RS4_HUMAN 40S RIBOSOMAL PROTEIN	2.17e+01
33	45	69.2	289	1	THTR_CHICK THIOSULFATE SULFURTRAN	2.17e+01
34	45	69.2	320	1	MEC3_CAEVU MECHANOSENSORY PROTEIN	2.17e+01
35	45	69.2	321	1	MEC3_CAEEL MECHANOSENSORY PROTEIN	2.17e+01
36	45	69.2	360	1	DCAM_SOLTU S-ADENOSYLMETHIONINE D	2.17e+01
37	45	69.2	361	1	DCAM_TOBAC S-ADENOSYLMETHIONINE D	2.17e+01
38	45	69.2	361	1	DCAM_NICSY S-ADENOSYLMETHIONINE D	2.17e+01
39	45	69.2	362	1	DCAM_DATST S-ADENOSYLMETHIONINE D	2.17e+01
40	45	69.2	474	1	CISY_EVENI CITRATE SYNTHASE, MITO	2.17e+01
41	45	69.2	475	1	CISY_ASNG CITRATE SYNTHASE, MITO	2.17e+01
42	45	69.2	880	1	VP2_ROTBR RNA-BINDING PROTEIN VP	2.17e+01
43	45	69.2	881	1	VP2_ROTBU RNA-BINDING PROTEIN VP	2.17e+01
44	45	69.2	881	1	VP2_ROTBU RNA-BINDING PROTEIN VP	2.17e+01
45	45	69.2	890	1	VP2_ROTWH RNA-BINDING PROTEIN VP	2.17e+01

ALIGNMENTS

RESULT	1	STANDARD;	PRT;	162 AA.
ID	OMP_HUMAN			
AC	P47874;			
DT	01-FEB-1996 (Rel. 33, Created)			
DT	01-FEB-1996 (Rel. 33, Last sequence update)			
DT	01-FEB-1996 (Rel. 33, Last annotation update)			
DE	OLFACATORY MARKER PROTEIN (OLFACATORY NEURONAL SPECIFIC PROTEIN).			
GN	OMP.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;			
OC	Eutheria; Primates; Catarrhini; Hominidae; Homo.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE: 9430732.			
RA	BULAKOVA O.I., RAMA KRISHNA N., GETCHELL T.V., MARGOLIS F.L.;			
RT	"Human and rodent OMP genes: conservation of structural and regulatory motifs and cellular localization.";			
RL	Genomics 20:452-462(1994).			
CC	-!- SUBCELLULAR LOCATION: CYTOPLASMIC.			
CC	-!- TISSUE SPECIFICITY: UNIQUELY ASSOCIATED WITH MATURE OLFACATORY RECEPTOR NEURONS.			
CC				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
DR	EMBL: U01212; AAA20485.1; -			
DR	MIM: 164340; -			
KW	Neurone; Olfaction; Acetylation.			
FT	INIT_MET 0 0 BY SIMILARITY.			
FT	MOD_RES 1 1 ACETYLATION (BY SIMILARITY).			
SQ	SEQUENCE 162 AA; 18805 MW; A0748827 CRC32;			

Query Match 81.5%; Score 53; DB 1; Length 162;  
Best Local Similarity 55.6%; Pred.No. 3.17e+01;  
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 46 QPAESVRL 54  
||:||||:  
QY 1 QPDDAVYKL 9

RESULT 2

ID ASEL\_YEAST STANDARD; PRT; 885 AA.  
AC P50275;  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 15-DEC-1999 (Rel. 39, Last annotation update)  
DE ANAPHASE SPINDLE ELONGATION PROTEIN.  
GN ASEL1 OR YOR038C OR YOR29-09.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;  
OC Saccharomycetaceae; Saccharomycetes.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA PELLMAN D., BAGGET M., TU Y.H., FINK G.R.;  
RT "Two microtubule-associated proteins required for anaphase spindle  
movement in Saccharomyces cerevisiae.";  
RL J. Cell Biol. 130:1373-1385(1995).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 97279235.  
RA VALENS M., BOHN C., DAIGNAN-FORNIER B., DANG V., BOLOTIN-FUKUHARA M.;  
RT "The sequence of a 54.7 kb fragment of yeast chromosome XV reveals  
the presence of two tRNAs and 24 new open reading frames.";  
RL Yeast 13:379-390(1997).  
CC -1- FUNCTION: REQUIRED FOR ANAPHASE SPINDLE ELONGATION.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; U20235; AAA75026.1; -;  
DR EMBL; 274566; CAA9251.1; -;  
DR EMBL; 270578; CAA94543.1; -;  
DR SGD; 10000125; ASEL1.  
KW Microtubules.  
SQ SEQUENCE 885 AA; 101623 MW; FF00B6B9 CRC32;  
  
Query Match 76.9%; Score 50; DB 1; Length 885;  
Best Local Similarity 44.4%; Pred. No. 1.65e+00;  
Matches 4; Conservative 4; Mismatches 1; Indels 0; Gaps 0;  
  
Db 823 EPEHSIYKL 831  
Qy 1 QPDDAVYKL 9  
  
RESULT 3  
ID RS4 CANAL STANDARD; PRT; 262 AA.  
AC P47837;  
DT 01-FEB-1996 (Rel. 33, Created)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE 40S RIBOSOMAL PROTEIN S4 (S7).  
GN RPS4 OR RPS7.  
OS Candida albicans (Yeast).  
OC Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;  
OC Candidaceae; Candida.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA DELBRUECK S., SONNEBORN A., GERADS M., GRABLOWITZ A.H., ERNST J.F.;  
RT "Characterization and regulation of the genes encoding ribosomal  
proteins L39 and S7 of the human pathogen Candida albicans.";  
RL Yeast 13:1199-1210(1997).  
CC -1- SIMILARITY: BELONGS TO THE S4E FAMILY OF RIBOSOMAL PROTEINS.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; X00412; CAA25118.1; -;  
DR EMBL; A02111; QRRRG.  
DR PIR; A28077; A28077.  
DR PFAM; PF00047; 1g; 5.  
KW Immunoglobulin domain; Repeat; Transmembrane; Glycoprotein; Signal;  
FT SIGNAL 1 18 POLYMERIC-IMMUNOGLOBULIN RECEPTOR.  
FT CHAIN 19 773 SECRETORY COMPONENT.  
FT CHAIN 19 615

CC use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; U37009; AAC49871.1; -;  
DR PROSITE; PS00528; RIBOSOMAL\_S4E; 1.  
DR PFAM; PF00900; Ribosomal\_S4e; 1.  
KW Ribosomal protein.  
SQ SEQUENCE 262 AA; 29204 MW; CE29056D CRC32;  
  
Query Match 75.4%; Score 49; DB 1; Length 262;  
Best Local Similarity 62.5%; Pred. No. 2.80e+00;  
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;  
  
Db 116 AEEAVYKL 123  
Qy 2 PDDAVYKL 9  
  
RESULT 4  
ID PIGR\_RABBIT STANDARD; PRT; 773 AA.  
AC P01832;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 21-JUL-1986 (Rel. 01, Last sequence update)  
DT 15-DEC-1999 (Rel. 39, Last annotation update)  
DE POLYMERIC-IMMUNOGLOBULIN RECEPTOR PRECURSOR (POLY-IG RECEPTOR) (PIGR)  
DE [CONTAINS: SECRETORY COMPONENT].  
GN PIGR.  
OS Oryctolagus cuniculus (Rabbit).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Lagomorpha; Leporidae; Oryctolagus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 84142246.  
RA MOSTOV K.E., FRIEDLANDER M., BLOBEL G.;  
RT "The receptor for transepithelial transport of Iga and Igm contains  
multiple immunoglobulin-like domains.";  
RL Nature 308:37-43(1984).  
RN [2]  
RP SEQUENCE OF 87-114 AND 410-428.  
RX MEDLINE; 88228032.  
RA FRUTIGER S., HUGHES G.J., HANLY W.C., JATON J.-C.;  
RT "Rabbit secretory components of different allotypes vary in their  
carbohydrate content and their sites of N-linked glycosylation.";  
RL J. Biol. Chem. 263:8120-8125(1988).  
CC -1- FUNCTION: THIS RECEPTOR BINDS POLYMERIC IGA AND IGM AT THE  
BASOLATERAL SURFACE OF EPITHELIAL CELLS. THE COMPLEX IS THEN  
TRANSPORTED ACROSS THE CELL TO BE SECRETED AT THE APICAL SURFACE.  
CC DURING THIS PROCESS A CLEAVAGE OCCURS THAT SEPARATE THE  
EXTRACELLULAR (KNOWN AS THE SECRETORY COMPONENT) FROM THE  
TRANSMEMBRANE SEGMENT.  
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. ALSO SECRETED.  
CC -1- POLYMORPHISM: THE SEQUENCE SHOWN IS THAT OF ALLOTYPIC T6.  
CC -1- SIMILARITY: CONTAINS 5 IMMUNOGLOBULIN-LIKE V-TYPE DOMAINS.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; X00412; CAA25118.1; -;  
DR EMBL; A02111; QRRRG.  
DR PIR; A28077; A28077.  
DR PFAM; PF00047; 1g; 5.  
KW Immunoglobulin domain; Repeat; Transmembrane; Glycoprotein; Signal;  
FT SIGNAL 1 18 POLYMERIC-IMMUNOGLOBULIN RECEPTOR.  
FT CHAIN 19 773 SECRETORY COMPONENT.  
FT CHAIN 19 615



FT DOMAIN 19 647 EXTRACELLULAR (POTENTIAL).  
FT TRANSFEM 648 670 POTENTIAL.  
FT DOMAIN 671 773 CYTOPLASMIC (POTENTIAL).  
FT DOMAIN 30 136 IG-LIKE V-TYPE DOMAIN 1.  
FT DOMAIN 137 243 IG-LIKE V-TYPE DOMAIN 2.  
FT DOMAIN 244 350 IG-LIKE V-TYPE DOMAIN 3.  
FT DOMAIN 351 456 IG-LIKE V-TYPE DOMAIN 4.  
FT DOMAIN 457 558 IG-LIKE V-TYPE DOMAIN 5.  
FT DISULFID 46 115 POTENTIAL.  
FT DISULFID 155 225 POTENTIAL.  
FT DISULFID 260 324 POTENTIAL.  
FT DISULFID 369 438 POTENTIAL.  
FT DISULFID 478 538 POTENTIAL.  
FT CARBOHYD 88 108 IN ALLOTYPES T61.  
FT CARBOHYD 108 108 IN ALLOTYPES T62 (PARTIAL) AND T63.  
FT CARBOHYD 418 418 K -> N (IN ALLOTYPES T61).  
FT VARIANT 88 88 D -> E (IN ALLOTYPES T61).  
FT VARIANT 94 94 TVDQLTON -> YLNRLSQS (IN ALLOTYPES T61).  
FT VARIANT 101 108 S -> T (IN ALLOTYPES T63).  
FT VARIANT 110 110  
SQ SEQUENCE 773 AA; 83886 MW; 79840D1F CRC32;

Query Match 75.4%; Score 49; DB 1; Length 773;  
Best Local Similarity 75.0%; Pred. No. 2.80e+00;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 137 EPDDVYK 144  
:|||||  
QY 1 QPDDAVYK 8

RESULT 5  
ID BUB1\_YEAST STANDARD; PRT; 1021 AA.  
AC P41695;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE CHECKPOINT SERINE/THRONINE-PROTEIN KINASE BUB1 (EC 2.7.1.-).  
GN BUB1 OR YGR188C OR G7542.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;  
OC Saccharomycetaceae; Saccharomycetes.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-S288C;  
RX MEDLINE; 95059057.  
RA ROBERTS B.T., FARR K.A., HOYT M.A.;  
RT "The Saccharomyces cerevisiae checkpoint gene BUB1 encodes a novel  
protein kinase.";  
RL Mol. Cell. Biol. 14:8282-8291(1994).  
[2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-S288C;  
RX MEDLINE; 97279231.  
RA ARROJO J., GARCIA-GONZALEZ M., GARCIA-SAEZ M.I., SANCHEZ-PEREZ M.,  
RA NOMBELA C.;  
RT "DNA sequence analysis of a 23,002 bp DNA fragment of the right arm  
of Saccharomyces cerevisiae chromosome VII.";  
RL Yeast 13:357-363(1997).  
CC -1- FUNCTION: INVOLVED IN CELL CYCLE CHECKPOINT ENFORCEMENT. CATALYZES  
CC THE PHOSPHORYLATION OF BUB3 AND ITS AUTOPHOSPHORYLATION.  
CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
CC -1- PTM: AUTOPHOSPHORYLATED.  
CC -1- SIMILARITY: WITH THE CONSERVED CATALYTIC DOMAINS OF SER/THR-  
CC PROTEIN KINASES.  
CC -1- SIMILARITY: SOME, IN THE N-TERMINUS WITH THE N-TERMINUS OF MAD3.  
CC -1- SIMILARITY: IN THE N-TERMINUS, WITH YEAST YJL013C.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>).

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC  
CC EMBL; L32027; AAA64894.1; -  
CC EMBL; Z72973; CAA97214.1; -  
CC EMBL; X99074; CAA67524.1; -  
CC SGD; L0000196; BUB1.  
CC DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
CC DR PROSITE; PS00108; PROTEIN\_KINASE\_ST; 1.  
CC DR PFAM; PF00069; kinase; 2.  
CC transferase; Serine/threonine-protein kinase; ATP-binding;  
KW Cell cycle; Phosphorylation; Nuclear protein.  
KW DOMAIN 705 1021 PROTEIN KINASE.  
FT NP\_BIND 711 719 ATP (BY SIMILARITY).  
FT BINDING 733 733 ATP.  
FT ACT\_SITE 833 833 BY SIMILARITY.  
FT MOTAGEN 733 733 K->R: LOSS OF ACTIVITY.  
FT CONFLICT 531 531 D -> V (IN REF. 1).  
SQ SEQUENCE 1021 AA; 117868 MW; C9532F44 CRC32;

Query Match 75.4%; Score 49; DB 1; Length 1021;  
Best Local Similarity 55.6%; Pred. No. 2.80e+00;  
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 300 QSNPNVYKL 308  
:|||||  
QY 1 QPDDAVYKL 9

RESULT 6  
ID TIAM\_MOUSE STANDARD; PRT; 1591 AA.  
AC Q60610;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE T-LYMPHOMA INVASION AND METASTASIS INDUCING PROTEIN 1 (TIAM1 PROTEIN).  
GN TIAM1 OR TIAM-1.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BALB/C; TISSUE-BRAIN;  
RX MEDLINE; 94243921.  
RA HABETS J.C., SCHOLTES E.H.M., ZUYDGEEST D., VAN DER KAMMEN R.A.,  
RA STAM J.C., BERNIS A., COLLARD J.G.;  
RT "Identification of an invasion-inducing gene, Tiam-1, that encodes a  
protein with homology to GDP-GTP exchangers for Rho-like proteins.";  
RL Cell 77:537-549(1994).  
CC -1- FUNCTION: MODULATES THE ACTIVITY OF RHO-LIKE PROTEINS AND CONNECTS  
CC EXTRACELLULAR SIGNALS TO CYTOSKELETAL ACTIVITIES. ACTS AS A GDP-  
CC DISSOCIATION STIMULATOR PROTEIN THAT STIMULATES THE GDP-GTP  
CC EXCHANGE ACTIVITY OF RHO-LIKE GTPASES AND ACTIVATES THEM.  
CC ACTIVATES RAC1, CDC42, AND TO A LESSER EXTENT RHOA (BY  
CC SIMILARITY). AFFECTS INVASIVENESS OF T-LYMPHOMA CELLS.  
CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN BRAIN AND TESTIS AND AT  
CC LOW OR MODERATE LEVELS IN ALMOST ALL OTHER NORMAL TISSUES. FOUND  
CC IN VIRTUALLY ALL ANALYZED TUMOR CELL LINES INCLUDING B- AND T-  
CC LYMPHOMAS, NEUROBLASTOMAS, MELANOMAS AND CARCINOMAS.  
CC -1- SIMILARITY: CONTAINS 1 DBL-HOMOLOGY DOMAIN (DH).  
CC -1- SIMILARITY: CONTAINS 2 PH DOMAINS.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; U05245; AAA18830.1; -  
CC SWISS-2DPAGE; Q60610; MOUSE.

DR MGD; MGI:103306; TIAM1.  
DR PROSITE; PS00741; GDS\_CDC24; 1.  
DR PROSITE; PS50003; PH\_DOMAIN; 1.  
DR PFAM; PF00169; PH; 1.  
DR PFAM; PF00595; PDZ; 1.  
DR PFAM; PF00621; RHOGEF; 1.  
KW Guanine-nucleotide releasing factor; Myristate.  
FT LIPID 2 MYRISTATE (POTENTIAL).  
FT DOMAIN 434 549 PH.  
FT DOMAIN 1047 1239 DH.  
FT DOMAIN 1261 1397 PH.  
FT DOMAIN 595 598 POLY-LYS.  
FT DOMAIN 1445 1449 POLY-ARG.  
SQ SEQUENCE 1591 AA; 177532 MW; 0220ECCC CRC32;  
  
Query Match 75.4%; Score 49; DB 1; Length 1591;  
Best Local Similarity 44.4%; Pred. No. 2.80e+00;  
Matches 4; Conservative 3; Mismatches 2; Indels 0; Gaps 0;  
  
Db 823 QPEDIYEL 831  
QY 1 QPDDAVYKL 9  
| | : : | |  
| | : : | |  
  
RESULT 7  
ID TIAM\_HUMAN STANDARD; PRT; 1591 AA.  
AC Q13009;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE T-LYMPHOMA INVASION AND METASTASIS INDUCING PROTEIN 1 (TIAM1 PROTEIN).  
OS TIAM1.  
GN Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Primates; Catarrhini; Homiidae; Homo.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=FETAL BRAIN;  
RX MEDLINE; 95249246.  
RA HABETS G.G.M., VAN DER KAMMEN R.A., STAM J.C., MICHIELS F.,  
COLLARD J.G.;  
RT "Sequence of the human invasion-inducing TIAM1 gene, its conservation  
in evolution and its expression in tumor cell lines of different  
tissue origin.";  
RL Oncogene 10:1371-1376(1995).  
RN [2]  
RP INTERACTIONS WITH RAC.  
RC TISSUE=BRAIN;  
RX MEDLINE; 95272708.  
RA MICHIELS F., HABETS G.G.M., STAM J.C., VAN DER KAMMEN R.A.,  
COLLARD J.G.;  
RT "A role for Rac in Tiam1-induced membrane ruffling and invasion.";  
RL Nature 375:338-340(1995).  
RN [3]  
RP MAPPING.  
RX MEDLINE; 95254877.  
RA HABETS G.G.M., VAN DER KAMMEN R.A., JENKINS N.A., GILBERT D.J.,  
COPELAND N.G., HAGEMELJER A., COLLARD J.G.;  
RT "The invasion-inducing TIAM1 gene maps to human chromosome band 21q22  
and mouse chromosome 16.";  
RL Cytogenet. Cell Genet. 70:48-51(1995).  
CC -!- FUNCTION: MODULATES THE ACTIVITY OF RHO-LIKE PROTEINS AND CONNECTS  
EXTRACELLULAR SIGNALS TO CYTOSKELETAL ACTIVITIES. ACTS AS A GDP-  
DISSOCIATION STIMULATOR PROTEIN THAT STIMULATES THE GDP-GTP  
EXCHANGE ACTIVITY OF RHO-LIKE GTPASES AND ACTIVATES THEM.  
CC ACTIVATES RAC1, CDC42, AND TO A LESSER EXTENT RHOA.  
CC -!- TISSUE SPECIFICITY: FOUND IN VIRTUALLY ALL ANALYZED TUMOR CELL  
LINES INCLUDING B- AND T-LYMPHOMAS, NEUROBLASTOMAS, MELANOMAS AND  
CARCINOMAS.  
CC -!- SIMILARITY: CONTAINS 1 DBL-HOMOLOGY DOMAIN (DH).  
CC -!- SIMILARITY: CONTAINS 2 PH DOMAINS.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>;  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL; Z68203; CAA92423.1; --  
DR EMBL; AE000101; AAB91896.1; --  
KW Hypothetical protein; Plasmid.  
SQ SEQUENCE 218 AA; 24594 MW; E40D26DE CRC32;  
  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>;  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL; U16296; AAA98443.1; --  
DR MIN; 600687; --  
DR PROSITE; PS00741; GDS\_CDC24; 1.  
DR PROSITE; PS50003; PH\_DOMAIN; 1.  
DR PFAM; PF00169; PH; 1.  
DR PFAM; PF00595; PDZ; 1.  
DR PFAM; PF00621; RHOGEF; 1.  
KW Guanine-nucleotide releasing factor; Myristate.  
FT LIPID 2 MYRISTATE (POTENTIAL).  
FT DOMAIN 434 549 PH.  
FT DOMAIN 1047 1239 DH.  
FT DOMAIN 1261 1397 PH.  
FT DOMAIN 595 598 POLY-LYS.  
FT DOMAIN 1445 1449 POLY-ARG.  
SQ SEQUENCE 1591 AA; 177637 MW; 759BC80E CRC32;  
  
Query Match 75.4%; Score 49; DB 1; Length 1591;  
Best Local Similarity 44.4%; Pred. No. 2.80e+00;  
Matches 4; Conservative 3; Mismatches 2; Indels 0; Gaps 0;  
  
Db 823 QPEDIYEL 831  
QY 1 QPDDAVYKL 9  
| | : : | |  
| | : : | |  
  
RESULT 8  
ID Y4VH\_RHISN STANDARD; PRT; 218 AA.  
AC Q53216;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE HYPOTHETICAL 24.6 KD PROTEIN Y4VH.  
GN Y4VH.  
OS Rhizobium sp. (strain NGR234).  
OG Plasmid sym pNGR234a.  
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
OC Rhizobiaceae; Rhizobium.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 97305956.  
RA FREIBERG C.A., FELLAY R., BARTOCH A., BROUGHTON W.J., ROSENTHAL A.,  
PERRET X.;  
RT "Molecular basis of symbiosis between Rhizobium and legumes.";  
RL Nature 387:394-401(1997).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 96389014.  
RA FREIBERG C., PERRET X., BROUGHTON W.J., ROSENTHAL A.;  
RT "Sequencing the 500-kb GC-rich symbiotic replicon of Rhizobium sp.  
NGR234 using dye terminators and a thermostable 'sequenase'; a  
beginning.";  
RL Genome Res. 6:590-600(1996).  
CC -!- SIMILARITY: NONE OBVIOUS.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>;  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL; Z68203; CAA92423.1; --  
DR EMBL; AE000101; AAB91896.1; --  
KW Hypothetical protein; Plasmid.  
SQ SEQUENCE 218 AA; 24594 MW; E40D26DE CRC32;

Query Match 73.8%; Score 48; DB 1; Length 218;  
Best Local Similarity 55.6%; Pred. No. 4.74e+00;  
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 188 HPNDPVYEL 196

Qy 1 QPDDAVYKL 9

RESULT 9

ID FAS3 DROME STANDARD; PRT; 508 AA.  
AC P15278;  
DT 01-APR-1990 (Rel. 14, Created)  
DT 01-APR-1990 (Rel. 14, Last sequence update)  
DT 01-JUN-1994 (Rel. 29, Last annotation update)  
DE FASCICLIN III PRECURSOR (FAS III).  
GN FAS3.

OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Insecta; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE; 90030406.

RA SNOW P.M., BIEBER A.J., GOODMAN C.S.;

RT "Fasciclin III: a novel homophilic adhesion molecule in Drosophila.";  
RL Cell 59:313-323(1989).

CC -1- FUNCTION: MEDIATES CELL ADHESION IN A CA-INDEPENDENT MANNER. IT  
DEVELOPING NERVOUS SYSTEM.

CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.

CC -1- TISSUE SPECIFICITY: EXPRESSED ON DIFFERENT SUBSETS OF AXON BUNDLES  
(FASCICLES) IN INSECT EMBRYOS.

CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. CONTAINS

ONE V-LIKE AND 2 C2-LIKE DOMAINS.

-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----  
CC EMBL; M27813; AAA28532.1; -.

CC PIR; A33378; A33378.

CC FLYBASE; FBgn0000636; Fas3.

CC Cell adhesion; Glycoprotein; Repeat; Immunoglobulin domain;

KW Transmembrane; Signal; Neurogenesis; Phosphorylation.

FT SIGNAL 1 20

FT CHAIN 21 508

FT DOMAIN 21 346

FT TRANSMEM 347 370

FT DOMAIN 371 508

FT DOMAIN 44 106

FT DOMAIN 143 218

FT DOMAIN 256 310

FT MOD\_RES 21 21

FT DISULFID 150 211

FT CARBOHYD 62 62

FT CARBOHYD 160 160

FT CARBOHYD 257 257

FT CARBOHYD 300 300

FT MOD\_RES 382 382

FT MOD\_RES 459 459

FT SEQUENCE 508 AA; 55883 MW; C9417AFB CRC32;

Query Match 73.8%; Score 48; DB 1; Length 508;  
Best Local Similarity 77.8%; Pred. No. 4.74e+00;

Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 240 QPDRAVYKL 248

Qy 1 QPDDAVYKL 9

RESULT 10

ID INR1 SHEEP STANDARD; PRT; 560 AA.

AC Q28589; Q95206;

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 15-JUL-1999 (Rel. 38, Last annotation update)

DE INTERFERON-ALPHA/BETA RECEPTOR ALPHA CHAIN PRECURSOR (IFN-ALPHA-REC)

DE (INTERFERON ALPHA/BETA RECEPTOR-1).

GN IFNARI OR IFNAR.

OS Ovis aries (Sheep).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

OC Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;

OC Caprinae; Ovis.

RN [1]

RP SEQUENCE FROM N.A.

RX TISSUE-ENDOMETRIUM;

RX MEDLINE; 97135690.

RA KALUZ S., FISHER P.A., KALUZOVA M., SHELDRIK E.L., FLINT A.P.F.;

RT "Structure of an ovine interferon receptor and its expression in

endometrium".

RL J. Mol. Endocrinol. 17:207-215(1996).

RN [2]

RP SEQUENCE FROM N.A.

RX TISSUE-ENDOMETRIUM;

RX MEDLINE; 98006426.

RA HAN C.-S., MATHALAGAN N., KLEMMAN S.W., ROBERTS R.M.;

RT "Molecular cloning of ovine and bovine type I interferon receptor

subunits from uteri, and endometrial expression of messenger

ribonucleic acid for ovine receptors during the estrous cycle and

pregnancy".

RL Endocrinology 138:4757-4767(1997).

CC -1- FUNCTION: RECEPTOR FOR INTERFERONS ALPHA AND BETA. BINDING TO TO

TYPE I IFNS TRIGGERS TYROSINE PHOSPHORYLATION OF A NUMBER OF

PROTEINS INCLUDING JAKS, TYK2, STAT PROTEINS AND IFN-R ALPHA- AND

BETA-SUBUNITS THEMSELVES.

CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.

CC -1- TISSUE SPECIFICITY: EXPRESSED IN ALL TISSUES EXAMINED EXCEPT

CONCEPTUS AT DAY 15 OF PREGNANCY.

CC -1- SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.

CC -1- SIMILARITY: BELONGS TO THE CLASS II CYTOKINE FAMILY OF RECEPTORS.

-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----  
CC EMBL; X95939; CAA65183.1; -.

CC EMBL; U65976; AB84231.1; -.

CC PFAM; PF00041; fn3; 1.

CC Receptor; Transmembrane; Glycoprotein; Signal.

CC BY SIMILARITY.

FT SIGNAL 1 24

FT CHAIN 25 560

FT DOMAIN 25 437

FT TRANSMEM 438 458

FT DOMAIN 459 560

FT DISULFID 76 84

FT DISULFID 199 220

FT CARBOHYD 47 47

FT CARBOHYD 55 55

FT CARBOHYD 85 85

FT CARBOHYD 108 108

FT CARBOHYD 109 109

FT CARBOHYD 172 172

FT CARBOHYD 222 222

FT CARBOHYD 285 285

```

FT CARBOHYD 47 47 POTENTIAL.
FT FT CARBOHYD 55 55 POTENTIAL.
FT FT CARBOHYD 85 85 POTENTIAL.
FT CARBOHYD 109 109 POTENTIAL.
FT FT CARBOHYD 172 172 POTENTIAL.
FT FT CARBOHYD 254 254 POTENTIAL.
FT CARBOHYD 313 313 POTENTIAL.
FT FT CARBOHYD 377 377 POTENTIAL.
FT FT CARBOHYD 434 434 POTENTIAL.
FT FT CONFLICT 422 422 F -> V (IN REF. 2).
SQ SEQUENCE 560 AA; 63818 MW; 44D98FDE CRC32;

Query Match 73.88; Score 48; DB 1; Length 560;
Best Local Similarity 62.98; Pred. No. 4.74e+00;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 185 PEDKIVKL 192
QY 2 PDDAVYKL 9
|:|:|:|

RESULT 12
ID OMP_RAT STANDARD; PRT; 162 AA.
AC P08523;
DC 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE OLFACTORY MARKER PROTEIN (OLFACTORY NEURONAL SPECIFIC PROTEIN).
GN OMP.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Cranial; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
[1]
RN RP SEQUENCE FROM N.A.
RX MEDLINE; 8717546.
RA ROGERS K.E.; DRAGUPTA P.; GUBLER U.; GRILLO M.; KHEW-GOODALL Y.S.;
RA MARGOLIS F.L.;
RT "Molecular cloning and sequencing of a cDNA for olfactory marker
RT protein.";
RL Proc. Natl. Acad. Sci. U.S.A. 84:1704-1708(1987).
[2]
RN RP SEQUENCE FROM N.A.
RX MEDLINE; 90046838.
RA DANCIGER E.; METTLING C.; VIDAL M.; MORRIS R.; MARGOLIS F.L.;
RT "Olfactory marker protein gene: its structure and olfactory neuron-
RT specific expression in transgenic mice.";
RL Proc. Natl. Acad. Sci. U.S.A. 86:8565-8569(1989).
[3]
RN RP SEQUENCE.
RX MEDLINE; 86321994.
RA SYDOR W.; TEITELBAUM Z.; BLACHER R.; SUN S.; BENZ W.; MARGOLIS F.L.;
RT "Amino acid sequence of a unique neuronal protein: rat olfactory
RT marker protein.";
RL Arch. Biochem. Biophys. 249:351-362(1986).
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- TISSUE SPECIFICITY: UNIQUELY ASSOCIATED WITH MATURE OLFACTORY
CC RECEPTOR NEURONS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC at the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M15644; AAA41757.1; -
CC EMBL; M36926; AAA03054.1; -
CC PIR; A27450; A27450.
CC PIR; A55025; A55025.
CC Neurone; Olfaction; Acetylation.
CC INIT_MET 0 0
CC MOD_RES 1 1 ACETYLATION.
FT FT

```

```

SQ SEQUENCE 162 AA; 18721 MW; 11FD9FB8 CRC32;
Query Match 72.3%; Score 47; DB 1; Length 162;
Best Local Similarity 44.4%; Pred. No. 7.94e+00;
Matches 4; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 46 RPAESVRL 54
:|:|:|:|
QY 1 QPDDAVYL 9

RESULT 13
ID G6PD_SCHPO STANDARD; PRT; 447 AA.
AC 000091;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1998 (Rel. 37, Last annotation update)
DE D-GLUCONO-6-PHOSPHATE 1-DEHYDROGENASE (EC 1.1.1.49) (G6PD) (FRAGMENT).
GN ZWF1 OR SPACJA12.18.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Archiascomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RA BADCOCK K., CHURCHER C.M., WOOD V., BARRELL B.G., RAJANDREAM M.A.;
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: D-GLUCOSE 6-PHOSPHATE + NADP(+) -
CC D-GLUCONO-DELTA-LACTONE 6-PHOSPHATE + NADPH.
CC -1- PATHWAY: FIRST STEP IN PENTOSE PHOSPHATE PATHWAY.
CC -1- SIMILARITY: BELONGS TO THE GLUCOSE-6-PHOSPHATE DEHYDROGENASE
CC FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z95395; CAB08746.1; -.
CC HSSP; P11411; IDPG.
CC PROSITE; PS00069; G6P_DEHYDROGENASE; 1.
CC PFAM; PF00479; G6PD; 1.
CC Oxidoreductase; NADP; Glucose metabolism.
KW Schizosaccharomycetes.
FT ACT_SITE 189 447 BY SIMILARITY.
FT NON_TER 447 447
SQ SEQUENCE 447 AA; 50926 MW; 59CB54F CRC32;

Query Match 72.3%; Score 47; DB 1; Length 447;
Best Local Similarity 71.4%; Pred. No. 7.94e+00;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 375 QPDEAY 381
:|:|:|:|
QY 1 QPDDAVY 7

RESULT 14
ID G6PD_CAEEL STANDARD; PRT; 522 AA.
AC Q27464;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE GLUCOSE-6-PHOSPHATE 1-DEHYDROGENASE (EC 1.1.1.49) (G6PD).
GN B0035.5.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
OC Rhabditina; Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.
RN [1]
RP SEQUENCE FROM N.A.

SQ SEQUENCE 522 AA; 60215 MW; D8B2BCE CRC32;

Query Match 72.3%; Score 47; DB 1; Length 522;
Best Local Similarity 62.5%; Pred. No. 7.94e+00;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 403 QPNEAYVM 410
:|:|:|:|
QY 1 QPDDAVYK 8

RESULT 15
ID YVFH_BACSU STANDARD; PRT; 563 AA.
AC P71067;
DT 01-NOV-1997 (Rel. 35, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE PUTATIVE L-LACTATE PERMEASE YVFH.
GN YVFH.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
RN [1]
RP SEQUENCE FROM N.A.
RA DENIZOT F.C.;
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 51-563 FROM N.A.
RC STRAIN=168;
RA FABRET C., QUENTIN Y., CHAPAL N., GUISEPPI A., HAIECH J., DENIZOT F.;
RL Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: MAY PLAY A ROLE IN L-LACTATE TRANSPORT.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE LLDP FAMILY OF TRANSPORTERS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z94043; CAB08002.1; -.
CC EMBL; Z71928; CAB96486.1; -.
CC EMBL; Z99121; CAB15424.1; -.
CC SUBTILIS; Bg11873; YVFH.
KW Hypothetical protein; Transport; Transmembrane.
FT TRANSMEM 14 34
POTENTIAL.

```

FT TRANSMEM 37 57 POTENTIAL.  
 FT TRANSMEM 73 93 POTENTIAL.  
 FT TRANSMEM 131 151 POTENTIAL.  
 FT TRANSMEM 157 177 POTENTIAL.  
 FT TRANSMEM 194 214 POTENTIAL.  
 FT TRANSMEM 220 240 POTENTIAL.  
 FT TRANSMEM 249 269 POTENTIAL.  
 FT TRANSMEM 304 324 POTENTIAL.  
 FT TRANSMEM 381 401 POTENTIAL.  
 FT TRANSMEM 419 439 POTENTIAL.  
 FT TRANSMEM 448 468 POTENTIAL.  
 FT TRANSMEM 506 526 POTENTIAL.  
 FT TRANSMEM 542 562 POTENTIAL.  
 SQ SEQUENCE 563 AA; 59761 MW; DB650DC4 CRC32;

Query Match 72.3%; Score 47; DB 1; Length 563;  
 Best Local Similarity 75.0%; Pred. No. 7.94e+00;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 371 PIEAVYKL 378  
 | :|||||  
 QY 2 PDDAVYKL 9

Search completed: Fri Apr 14 23:38:29 2000  
 Job time : 48 secs.

\*\*\*\*\*  
[M][A][P][S][R][C][H] (TM)  
\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Fri Apr 14 23:38:47 2000; MasPar time 12.88 Seconds  
48.442 Million cell updates/sec  
Tabular output not generated.

Title: >US-08-452-843-7  
Description: (1-9) from US08452843.pep  
Perfect Score: 65  
Sequence: 1 QPDDAVYKL 9

Scoring table: PAM 150  
Gap 15

Searched: 225878 seqs, 69334122 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: sptrembl12  
1:sp\_archaea 2:sp\_bacteria 3:sp\_fungi 4:sp\_human  
5:sp\_invertebrate 6:sp\_mammal 7:sp\_mhc 8:sp\_organelle  
9:sp\_phase 10:sp\_plant 11:sp\_rodent 12:sp\_unclassified  
13:sp\_vertebrate 14:sp\_virus

Statistics: Mean 22.991; Variance 25.213; scale 0.912

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	51	78.5	347	1	IMMUNOGENIC PROTEIN (B	2.19e+00
2	51	78.5	437	4	HYPOTHETICAL 48.5 KD P	2.19e+00
3	50	76.9	259	1	259AA LONG HYPOTHETICA	3.73e+00
4	50	76.9	771	4	COLLAPIN-1 PRECURSOR	3.73e+00
5	50	76.9	772	11	COLLAPIN-1 PRECURSOR	3.73e+00
6	50	76.9	772	11	COLLAPIN-1 PRECURSOR	3.73e+00
7	50	76.9	1213	1	MANNOSE-6-PHOSPHATE	3.73e+00
8	49	75.4	832	2	HYPOTHETICAL 92.9 KD P	6.28e+00
9	48	73.8	1307	5	T22C1.10 PROTEIN.	1.05e+01
10	47	72.3	163	11	OLFACTORY MARKER PROTE	1.75e+01
11	47	72.3	317	10	GLUCOSE-6-PHOSPHATE 1-	1.75e+01
12	47	72.3	465	10	GLUCOSE-6-PHOSPHATE 1-	1.75e+01
13	47	72.3	492	10	GLUCOSE-6-PHOSPHATE 1-	1.75e+01
14	47	72.3	509	11	CYCLOCHROME P-450.	1.75e+01
15	47	72.3	588	10	GLUCOSE-6-PHOSPHATE 1-	1.75e+01
16	47	72.3	599	10	GLUCOSE-6-PHOSPHATE 1-	1.75e+01
17	47	72.3	604	10	GLUCOSE-6-PHOSPHATE 1-	1.75e+01
18	47	72.3	689	4	KIAA0846 PROTEIN.	1.75e+01
19	47	72.3	895	5	GAG PROTEIN.	1.75e+01
20	47	72.3	1215	2	DNA FOR SEROTYPE B CAP	1.75e+01

21	46	70.8	198	10	Q92R37	DSPTP1 PROTEIN.	2.88e+01
22	46	70.8	254	2	O67279	DMSO REDUCTASE CHAIN B	2.88e+01
23	46	70.8	289	3	O59783	HYPOTHETICAL 33.3 KD P	2.88e+01
24	46	70.8	387	2	P73562	CARBOXYNORSPERMIDINE D	2.88e+01
25	46	70.8	414	2	O9WZ48	CONSERVED HYPOTHETICAL	2.88e+01
26	46	70.8	424	2	O47628	TIEB PROTEIN.	2.88e+01
27	46	70.8	515	11	O55044	GLUCOSE-6-PHOSPHATE 1-	2.88e+01
28	46	70.8	689	13	O13265	TISSUE TRANSGLUTAMINAS	2.88e+01
29	46	70.8	787	2	O67027	NADH DEHYDROGENASE I C	2.88e+01
30	46	70.8	1099	2	O9WZ27	CARBAMOYL-PHOSPHATE SY	2.88e+01
31	46	70.8	1392	10	O82493	T12H20.12 PROTEIN.	2.88e+01
32	45	69.2	176	5	P90669	CARBOXYPEPTIDASE-RELAT	4.70e+01
33	45	69.2	289	5	P91237	COSMID F08D12.	4.70e+01
34	45	69.2	346	5	O02240	MEC-3 PROTEIN.	4.70e+01
35	45	69.2	508	2	P73375	L-THREONINE DEAMINASE.	4.70e+01
36	45	69.2	580	2	O87374	DIGUANYLATE CYCLASE.	4.70e+01
37	45	69.2	583	2	O32748	PLASMID DNA FOR HORA.	4.70e+01
38	45	69.2	605	2	O84668	DNA GYRASE SUBUNIT B.	4.70e+01
39	45	69.2	660	2	O9WXR2	OLIGOPEPTIDE ABC TRANS	4.70e+01
40	45	69.2	882	14	O86218	CAPSID PROTEIN VP2.	4.70e+01
41	45	69.2	882	14	O89813	VP2 GENOMIC RNA, COMPL	4.70e+01
42	45	69.2	897	14	O55591	RNA FOR VP2, COMPLETE	4.70e+01
43	45	69.2	1446	5	O77063	CARBOXYPEPTIDASE D.	4.70e+01
44	44	67.7	203	2	O30343	HEMAGGLUTININ/PROTEASE	7.61e+01
45	44	67.7	1083	10	O48839	PUTATIVE UBIQUITIN SPE	7.61e+01

ALIGNMENTS

RESULT	1	PRELIMINARY;	PRT;	347 AA.
ID	O29274			
AC	O29274;			
DT	01-JAN-1998 (Tremblrel. 05, Created)			
DT	01-JAN-1998 (Tremblrel. 05, Last sequence update)			
DT	01-AUG-1998 (Tremblrel. 07, Last annotation update)			
DE	IMMUNOGENIC PROTEIN (BCSP31-3).			
GN	Af0988.			
OS	Archaeoglobus fulgidus.			
OC	Archaea; Euryarchaeota; Archaeoglobales; Archaeoglobaceae;			
OC	Archaeoglobus.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=VC-16 / DSM 4304 / ATCC 49558;			
RX	MEDLINE; 98049343.			
RA	KLENK H.-F., CLAYTON R.A., TOMB J.-F., WHITE O., NELSON K.E.,			
RA	KETCHUM K.A., DODSON R.J., GWINN M., HICKEY E.K., PETERSON J.D.,			
RA	RICHARDSON D.L., KERLAVAGE A.R., GRAHAM D.E., KYRPIDES N.C.,			
RA	FLEISCHMANN R.D., QUACKENBUSH J., LEE N.H., SUTTON G.G., GILL S.,			
RA	KIRKNESS E.F., DOUGHERTY B.A., MCKENNEY K., ADAMS M.D., LOFTUS B.,			
RA	PETERSON S., REICH C.I., MCNEIL L.K., BADGER J.H., GLODEK A., ZHOU L.,			
RA	OVERBECK R., GOCCAYNE J.D., WEIDMAN J.F., McDONALD L., UTTERBACK T.,			
RA	COTTON M.D., SPRIGGS T., ARTIACH P., KAINE B.P., SYKES S.M.,			
RA	SADOM P.W., D'ANDREA K.P., BOWMAN C., FUJII C., GARLAND S.A.,			
RA	MASON T.M., OLSEN G.J., FRASER C.M., SMITH H.O., WOESE C.R.,			
RA	VENTER J.C.;			
RT	"The complete genome sequence of the hyperthermophilic, sulphate-			
RT	reducing archaeon Archaeoglobus fulgidus."			
RL	Nature 390:364-370(1997).			
DR	EMBL; AF001036; AAB90255.1; -			
DR	TIGR; AF0988; -			
KW	Hypothetical protein.			
SW	SEQUENCE 347 AA; 764D267E CRC32;			

Query Match 78.5%; Score 51; DB 1; Length 347;  
Best Local Similarity 75.0%; Pred. No. 2.19e+00;  
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 273 PEDAVYNL 280

QY 2 PDADVYKL 9

RESULT 2

```

ID Q99764 PRELIMINARY; PRT; 437 AA.
AC Q99764;
DT 01-MAY-1997 (TEMBLrel. 03, Created)
DT 01-MAY-1997 (TEMBLrel. 03, Last sequence update)
DT 01-NOV-1999 (TEMBLrel. 12, Last annotation update)
DE HYPOTHETICAL 48.5 KD PROTEIN (FRAGMENT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN;
RX MEDLINE; 96207227.
RA ANDERSSON B., WENTLAND M.A., RICAFFENTE J.Y., LIU W., GIBBS R.A.;
RT "A 'double adaptor' method for improved shotgun library
RL Anal. Biochem. 236:107-113(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN;
RA YU W., ANDERSSON B., WORLEY K.C., MUZY D.M., DING Y., LIU W.,
RA RICAFFENTE J.Y., WENTLAND M.A., LENNON G., GIBBS R.A.;
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; U79241; AABSO199.1; -.
DR PFAM; PFO1507; PAPS_reduct; 1.
KW Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 437 AA; 48483 MW; 4590C34B CRC32;

Query Match 78.5%; Score 51; DB 4; Length 437;
Best Local Similarity 66.7%; Pred. No. 2.19e+00;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 211 QASEAVYKL 219
QY 1 QPDDAVYKL 9

RESULT 3
ID Q9VAC3 PRELIMINARY; PRT; 259 AA.
AC Q9VAC3;
DT 01-NOV-1999 (TEMBLrel. 12, Created)
DT 01-NOV-1999 (TEMBLrel. 12, Last sequence update)
DT 01-NOV-1999 (TEMBLrel. 12, Last annotation update)
DE 259AA LONG HYPOTHETICAL PROTEIN.
GN APE2016.
OS Aeropyrum pernix.
OC Archaea; Crenarchaeota; Aeropyrum.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K1;
RX MEDLINE; 99310339.
RA KAWABAYASHI Y., HINO Y., HORIKAWA H., YAMAZAKI S., HAIKAWA Y.,
RA JIN-NO K., TAKAHASHI M., SEKINE M., BABA S., ANKAI A., KOSUGI H.,
RA HOSOVAMA A., FUKUI S., NAGAI Y., NISHIJIMA K., NAKAZAWA H.,
RA TAKAMIYA M., MASUDA S., FUNAHASHI T., TANAKA T., KUDOH Y.,
RA YAMAZAKI J., KUSHIDA N., OGUCHI A., AOIKI K., KUBOTA K., NAKAMURA Y.,
RA NOMURA N., SAKO Y., KIKUCHI H.;
RT "Complete genome sequence of an aerobic hyper-thermophilic
RCrenarchaeon, Aeropyrum pernix K1."
RL DNA Res. 6:83-101(1999).
DR EMBL; AP000063; BAA81026.1; -.
SQ SEQUENCE 259 AA; 28810 MW; 1CE2AED1 CRC32;

Query Match 76.9%; Score 50; DB 1; Length 259;
Best Local Similarity 55.6%; Pred. No. 3.73e+00;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 149 EPEDPVYTL 157
QY 1 QPDDAVYKL 9

```

```

RESULT 4
ID Q14563 PRELIMINARY; PRT; 771 AA.
AC Q14563;
DT 01-NOV-1996 (TEMBLrel. 01, Created)
DT 01-NOV-1996 (TEMBLrel. 01, Last sequence update)
DT 01-NOV-1999 (TEMBLrel. 12, Last annotation update)
DE COLLAPLIN-1 PRECURSOR (SEMAPHORIN III) (SEMAPHORIN D).
GN HSEMA-III.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN;
RX MEDLINE; 94094332.
RA KOLODKIN A.L., MATTHES D.J., GOODMAN C.S.;
RT "The semaphorin genes encode a family of transmembrane and secreted
RT growth cone guidance molecules."
RL Cell 75:1389-1399(1993).
RN [2]
RP SEQUENCE OF 1-37 FROM N.A.
RA WOESSNER J., MINX P., HINDS K., STROMWATT C.;
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 1-37 FROM N.A.
RA WATERSTON R.;
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE OF 39-182 FROM N.A.
RA ROHLFING T., TIN-WOLLAM A.M., DUCKELS G.;
RT "The sequence of Homo sapiens PAC clone DJ0649P17."
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE OF 39-182 FROM N.A.
RA WATERSTON R.H.;
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
RN [6]
RP SEQUENCE OF 39-182 FROM N.A.
RA WATERSTON R.;
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: INDUCES THE COLLAPSE AND PARALYSIS OF NEURONAL GROWTH
CC CONES. COULD SERVE AS A LIGAND THAT GUIDES SPECIFIC GROWTH CONES
CC BY A MOTILITY-INHIBITING MECHANISM (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: SECRETED (BY SIMILARITY).
CC -1- DOMAIN: THE C-TERMINAL HALF OF COLLAPLIN CONTAINS A SINGLE
CC IMMUNOGLOBULIN-LIKE DOMAIN AND AN ADDITIONAL HIGHLY BASIC REGION.
CC THE N-TERMINAL HALF OF COLLAPLIN SHARES SIGNIFICANT HOMOLOGY WITH
CC FASCICLIN IV. STRONG BINDING TO NEUROPILIN IS MEDIATED BY THE
CC CARBOXY THIRD OF THE COLLAPLIN.
CC -1- PTM: PROCESSED BY A URIN-LIKE ENDOPEPTIDASE. THIS PROCESSING
CC ACTIVATES ITS REPULSIVE PROPERTIES, GENERATES FUNCTIONALLY
CC DIFFERENT ISOFORMS, AND IS COUPLED TO AN ADDITIONAL STEP FOR
CC ACTIVATION (BY SIMILARITY).
CC -1- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAIN.
CC -1- SIMILARITY: BELONGS TO THE SEMAPHORIN FAMILY.
DR EMBL; L26081; AAB65938.1; -.
DR EMBL; AC004451; AAC06185.1; -.
DR EMBL; AC004848; AAC78622.1; -.
DR PFAM; PF000047; ig; 1.
DR PFAM; PF01403; Sema; 1.
KW Immunoglobulin domain; Signal; Glycoprotein;
KW Cleavage on pair of basic residues.
FT SIGNAL 1 20
FT CHAIN 21 771
FT DOMAIN 240 538
FT DOMAIN 642 728
FT DOMAIN 727 769
FT DISULFID 649 722
FT CARBOHYD 53 53
FT CARBOHYD 125 125
FT CARBOHYD 590 590
FT SEQUENCE 771 AA; 88889 MW; 9EB1A137 CRC32;

```





```
FT CONFLICT 620 620 E -> R (IN REF. 2).
FT CONFLICT 623 623 R -> K (IN REF. 3).
FT CARBOHYD 53 53 POTENTIAL.
FT CARBOHYD 125 125 POTENTIAL.
FT CARBOHYD 591 591 POTENTIAL.
SQ SEQUENCE 772 AA; 88799 MW; 4F0698CF CRC32;

Query Match 76.9%; Score 50; DB 11; Length 772;
Best Local Similarity 44.4%; Pred. No. 3.73e+00;
Matches 4; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 149 HPDENIFKL 157
:|:|:|:|
QY 1 QPDDAVYKL 9

RESULT 7 PRELIMINARY; PRT; 1213 AA.
ID O30191
AC O30191;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE MANNOSYLTRANSFERASE A (MTFA).
GN AF0045.
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobales; Archaeoglobaceae;
OC Archaeoglobus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE; 98049343.
RA KLENK H.-P., CLAYTON R.A., TOMB J.-F., WHITE O., NELSON K.E.,
RA KETCHUM K.A., DODSON R.J., GWINN M., HICKEY E.K., PETERSON J.D.,
RA RICHARDSON D.L., KERLAVAGE A.R., GILL S.,
RA FLEISCHMANN R.D., QUACKENBUSH J., LEE N.H., SUTTON G.G., GILL S.,
RA KIRKNESS E.F., DOUGHERTY B.A., MCKENNEY K., ADAMS M.D., LOFTUS B.,
RA PETERSON S., REICH C.I., MCNEIL L.K., BADGER J.H., GLODEK A., ZHOU L.,
RA OVERBEER R., GOCAYNE J.D., WEIDMAN J.F., McDONALD L., UTERBACK T.,
RA COTTON M.D., SPRIGGS T., ARTTACH P., KATZ B.P., SYKES S.M.,
RA SADOW P.W., D'ANDREA K.P., BOWMAN C., FUJII C., GARLAND S.A.,
RA MASON T.M., OLSEN G.J., FRASER C.M., SMITH H.O., WOESE C.R.,
RA VENTER J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
reducing archaeon Archaeoglobus fulgidus."
RL Nature 390:364-370(1997).
DR EMBL; AE001103; AAB91182.1; -.
DR TIGR; AF0045; -.
DR PFAM; PF00534; Glycos_transf_1; 3.
KW Hypothetical protein; Transferase; Glycosyltransferase.
SQ SEQUENCE 1213 AA; 140592 MW; B63A3D1F CRC32;

Query Match 76.9%; Score 50; DB 1; Length 1213;
Best Local Similarity 62.5%; Pred. No. 3.73e+00;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 730 PNEVYKL 737
:|:|:|:|
QY 2 PDDAVYKL 9

RESULT 8 PRELIMINARY; PRT; 832 AA.
ID P74619
AC P74619;
DT 01-FEB-1997 (TrEMBLrel. 02, Created)
DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT 01-JAN-1999 (TrEMBLrel. 09, Last annotation update)
DE HYPOTHETICAL 92.9 KD PROTEIN.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-PCC6803;
RA TABATA S.;

Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
[2]
RN SEQUENCE FROM N.A.
RC STRAIN-PCC6803;
RX MEDLINE; 97061201.
RA KANEKO T., SATO S., KOTANI H., TANAKA A., ASAMIZU E., NAKAMURA Y.,
RA MIYAJIMA N., HIROSAWA M., SUGIURA M., SASAMOTO S., KIMURA T.,
RA HOSOUCHI T., MATSUNO A., MURAKI A., NAKAZAKI N., NARUO K., OKUMURA S.,
RA SHIMO S., TAKEUCHI C., WADA T., WATANABE A., YAMADA M., YASUDA M.,
RA TABATA S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions.";
RL DNA Res. 3:109-136(1996).
DR EMBL; D90916; BAA18727.1; -.
KW Hypothetical protein.
SQ SEQUENCE 832 AA; 92865 MW; 08554B2A CRC32;

Query Match 75.4%; Score 49; DB 2; Length 832;
Best Local Similarity 44.4%; Pred. No. 6.28e+00;
Matches 4; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 309 QPDEQIFRL 317
:|:|:|:|
QY 1 QPDDAVYKL 9

RESULT 9 PRELIMINARY; PRT; 1307 AA.
ID Q22670
AC Q22670;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JAN-1999 (TrEMBLrel. 09, Last annotation update)
DE T22C1.10 PROTEIN.
GN T22C1.10.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;
OC Rhabditina; Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.
RN [1]
RP SEQUENCE FROM N.A.
RA MCMURRAY A.;
RT Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
[2]
RN SEQUENCE FROM N.A.
RX MEDLINE; 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M., COULSON A.,
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., FULTON L.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIEKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SWALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WORLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
DR EMBL; 275550; CAA99926.1; -.
SQ SEQUENCE 1307 AA; 149908 MW; 54514C24 CRC32;

Query Match 73.8%; Score 48; DB 5; Length 1307;
Best Local Similarity 62.5%; Pred. No. 1.05e+01;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 118 PDEPVYKI 125
:|:|:|:|
QY 2 PDDAVYKL 9

RESULT 10 PRELIMINARY; PRT; 163 AA.
ID Q64288
AC Q64288;
```

```

Db 202 QPDEATY 208
    |||:|:|
QY 1 QPDDAVY 7

RESULT 12
ID Q24359 PRELIMINARY: PRT; 465 AA.
AC Q24359;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
DE GLUCOSE-6-PHOSPHATE 1-DEHYDROGENASE (EC 1.1.1.49) (G6PD) (FRAGMENT).
DE G6PD.
GN Spinacia oleracea (Spinach).
OS Spinacia oleracea (Spinach).
OG Plasmid pZL1.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Caryophyllidae; Caryophyllales; Chenopodiaceae;
OC Spinacia.
OC [1]
RN SEQUENCE FROM N.A.
RP STRAIN=CV. MATADOR; TISSUE=LEAVES;
RC FUNK A., DIOGNI T., PERROUD P.F., CRESPI P., GREPPIN H.;
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: D-GLUCOSE 6-PHOSPHATE + NADP(+) = D-GLUCONIC
CC -1- DELTA-LACTONE 6-PHOSPHATE + NADPH.
CC -1- PATHWAY: FIRST STEP IN PENTOSE PHOSPHATE PATHWAY.
CC EMBL; AJ000184; CAA03941.1; -.
DR HSSP; P11411; 2DPG.
DR MENDEL; 26977; Sp101; 2614; 26977.
DR PROSITE; PS00069; G6P_DEHYDROGENASE; 1.
DR PFAM; PF00479; G6PD; 1.
KW Oxidoreductase; Plasmid; NADP; Glucose metabolism.
FT NON_TER 1
FT ACT_SITE 225 225 BY SIMILARITY.
FT NON_TER 465 465
FT NON_TER 465 465
SQ SEQUENCE 465 AA; 53101 MW; 929DA6AC CRC32;

Query Match 72.3%; Score 47; DB 10; Length 465;
Best Local similarity 71.4%; Pred. No. 1.75e+01;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps

Db 420 QPDEATY 426
    |||:|:|
QY 1 QPDDAVY 7

RESULT 13
ID Q43728 PRELIMINARY: PRT; 492 AA.
AC Q43728;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
DE GLUCOSE-6-PHOSPHATE 1-DEHYDROGENASE (EC 1.1.1.49) (G6PD) (FRAGMENT).
DE G6PD.
GN Arabidopsis thaliana (Mouse-ear cress).
OS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
OC Arabidopsis.
OC [1]
RN SEQUENCE FROM N.A.
RP STRAIN=CV. COLUMBIA; TISSUE=LEAVES, STEMS, SOME FLOWERS, AND ROOTS;
RC FUNK A., GREPPIN H., TACCINI P.;
RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: D-GLUCOSE 6-PHOSPHATE + NADP(+) = D-GLUCONIC
CC -1- DELTA-LACTONE 6-PHOSPHATE + NADPH.
CC -1- PATHWAY: FIRST STEP IN PENTOSE PHOSPHATE PATHWAY.
CC EMBL; X84229; CAA59011.1; -.
DR HSSP; P11411; 2DPG.
DR MENDEL; 16916; Arath; 2614; 16916.
DR PROSITE; PS00069; G6P_DEHYDROGENASE; 1.
DR PFAM; PF00479; G6PD; 1.

```

KW Oxidoreductase; NADP; Glucose metabolism.  
FT NON\_TER 1  
FT ACT\_SITE 183 183 BY SIMILARITY.  
SQ SEQUENCE 492 AA; 56203 MW; 8D214E1E CRC32;  
Query Match 72.3%; Score 47; DB 10; Length 492;  
Best Local Similarity 71.4%; Pred. No. 1.75e+01;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
Db 378 QPDEAIY 384  
|||:|:|  
QY 1 QPDDAVY 7

RESULT 14  
ID O88933 PRELIMINARY; PRT; 509 AA.  
AC O88933;  
DT 01-NOV-1998 (TREMBLrel. 08, Created)  
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)  
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)  
DE CYTOCHROME P-450.  
GN CYP4A10  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-DDY; TISSUE=LIVER;  
RA YASUMURA N., IKEDA T.;  
RT "Polymorphism of cyp 4A10 sequence between C57BL/6 and ddv mouse."  
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB018421; BAA33804.1;  
DR PFAM; PF00067; p450; 1.  
DR PRINTS; PR00385; P450.  
DR PRINTS; PR00464; EP450II.  
SQ SEQUENCE 509 AA; 58339 MW; 29A7213C CRC32;

Query Match 72.3%; Score 47; DB 11; Length 509;  
Best Local Similarity 44.4%; Pred. No. 1.75e+01;  
Matches 4; Conservative 4; Mismatches 1; Indels 0; Gaps 0;  
Db 236 HQNDTIYKL 244  
:|:|:|  
QY 1 QPDDAVYKL 9

RESULT 15  
ID O65856 PRELIMINARY; PRT; 588 AA.  
AC O65856;  
DT 01-AUG-1998 (TREMBLrel. 07, Created)  
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)  
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)  
DE GLUCOSE-6-PHOSPHATE 1-DEHYDROGENASE (EC 1.1.1.49) (G6PD) PRECURSOR.  
OS Nicotiana tabacum (Common tobacco).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;  
OC core eudicots; Asteridae; euasterids I; Solanales; Solanaceae;  
OC Nicotiana.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV. SAMSUN NN; TISSUE=YOUNG SINK LEAVES;  
RA HAUSCHILD R., LANGE C., PIETERSMA M., WENDT U., VONSCHEWEN A.;  
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.  
CC -1- CATALYTIC ACTIVITY: D-GLUCOSE 6-PHOSPHATE + NADP(+) -> D-GLUCONO-  
CC DELTA-LACTONE 6-PHOSPHATE + NADPH.  
CC -1- PATHWAY: FIRST STEP IN PENTOSE PHOSPHATE PATHWAY.  
DR EMBL; AJ001772; CAA04994.1; -;  
DR HSSP; P11411; 2DPG.  
DR MENDEL; 29933; N1cta;2614;29933.  
DR PROSITE; PS00069; G6P\_DEHYDROGENASE; 1.  
DR PFAM; PF00479; G6PD; 1.  
DR PRINTS; PR00079; G6PDHGRNASE.  
KW Signal; Oxidoreductase; NADP; Glucose metabolism.

FT SIGNAL 1 66 POTENTIAL.  
FT ACT\_SITE 279 279 BY SIMILARITY.  
SQ SEQUENCE 588 AA; 66801 MW; EDA5A1F0 CRC32;  
Query Match 72.3%; Score 47; DB 10; Length 588;  
Best Local Similarity 71.4%; Pred. No. 1.75e+01;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
Db 474 QPDEAIY 480  
|||:|:|  
QY 1 QPDDAVY 7

Search completed: Fri Apr 14 23:40:32 2000  
Job time : 105 secs.

\*\*\*\*\*  
W P S R L H  
\*\*\*\*\*  
(TM)  
\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Apr 14 23:42:20 2000; Maspar time 4.79 Seconds  
Tabular output not generated. 44.470 Million cell updates/sec

Title: >US-08-452-843-8  
Description: (1-9) from US08452843.pep  
Perfect Score: 73  
Sequence: 1 IPYPIVRKL 9

Scoring table: PAM 150  
Gap 15

Searched: 189963 seqs, 23686106 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: a-geneseq36  
1:geneseq36

Statistics: Mean 17.592; Variance 47.836; scale 0.368

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	73	100.0	9	1 R89369	Cw6 consensus peptide	1.23e-01
2	65	89.0	9	1 R89370	Cw6 consensus peptide	1.27e+00
3	53	72.6	9	1 R89371	Cw6 consensus peptide	3.66e+01
4	51	69.9	133	1 W28511	Product of clone L105	6.26e+01
5	51	69.9	133	1 W50884	Amino acid sequence of	6.26e+01
6	50	68.5	356	1 P70388	D-amino acid oxidase	8.17e+01
7	50	68.5	356	1 R04066	T-variabilis D-amino a	8.17e+01
8	49	67.1	1199	1 W47206	Bos taurus tubulin-fo1	1.06e+02
9	49	67.1	1676	1 R77604	Pro-C5 polypeptide.	1.06e+02
10	48	65.8	159	1 W44125	Streptococcus pneumoni	1.38e+02
11	48	65.8	311	1 W25084	Haemophilus influenzae	1.38e+02
12	48	65.8	833	1 W32114	Streptococcus pneumoni	1.38e+02
13	47	64.4	216	1 W73419	Human secreted protein	1.80e+02
14	47	64.4	319	1 W53896	Human G-protein couple	1.80e+02
15	47	64.4	319	1 W69735	Human C5a-like protein	1.80e+02
16	47	64.4	319	1 W52991	Homo sapiens clone H96	1.80e+02
17	47	64.4	458	1 W63740	HIV-1 NL-43 gag protei	1.80e+02
18	46	63.0	245	1 W08080	Bovine oncostatin M.	2.33e+02
19	46	63.0	321	1 W53243	Mus musculus vascular	2.33e+02
20	46	63.0	325	1 W53240	Homo sapiens vascular	2.33e+02
21	46	63.0	326	1 W44296	Rat vascular endotheli	2.33e+02
22	46	63.0	354	1 W53241	Homo sapiens vascular	2.33e+02
23	46	63.0	354	1 W49036	Human zveg12 growth fa	2.33e+02

24	46	63.0	354	1 W44293	Human vascular endothe	2.33e+02
25	46	63.0	358	1 W53242	Mus musculus vascular	2.33e+02
26	46	63.0	358	1 W4295	Mouse vascular endothe	2.33e+02
27	46	63.0	358	1 W14992	Murine c-Fos induced g	2.33e+02
28	46	63.0	592	1 R96247	Malic enzyme #2	2.33e+02
29	46	63.0	620	1 W14594	Human c-Fos induced gr	2.33e+02
30	46	63.0	736	1 Y07046	Breast cancer associat	2.33e+02
31	46	63.0	749	1 P70286	Protein encoded by pla	2.33e+02
32	46	63.0	795	1 W97842	Human p2y11 receptor.	2.33e+02
33	46	63.0	1330	1 R15444	Swine herpes virus-1 m	2.33e+02
34	46	63.0	251	1 W37358	HISF protein involved	3.01e+02
35	46	63.0	476	1 W06782	ILRV protein kinase.	3.01e+02
36	46	63.0	585	1 R96246	Malic enzyme #1.	3.01e+02
37	46	63.0	1635	1 W34624	Human C3 protein mutan	3.01e+02
38	46	63.0	1661	1 W34625	Human C3 protein mutan	3.01e+02
39	46	63.0	1663	1 W34614	Human C3 protein mutan	3.01e+02
40	46	63.0	1663	1 W34609	Human C3 protein mutan	3.01e+02
41	46	63.0	1663	1 W34610	Human C3 protein mutan	3.01e+02
42	46	63.0	1663	1 W40989	Human C3 protein mutan	3.01e+02
43	46	63.0	1663	1 W40990	Human C3 protein mutan	3.01e+02
44	46	63.0	1663	1 W34620	Human C3 protein mutan	3.01e+02
45	46	63.0	1663	1 W34617	Human C3 protein mutan	3.01e+02

ALIGNMENTS

RESULT 1  
ID R89369 standard; peptide; 9 AA.  
AC R89369;  
DT 18-SEP-1996 (first entry)  
DE Cw6 consensus peptide derived immunogenic peptide #1.  
KW Immunogenic peptide; supermotif; HLA molecule; CTL response;  
KW therapeutic; diagnostic; cancer; viral infection; hepatitis B;  
KW hepatitis C.  
OS Synthetic.  
PN W09603140-A1.  
PD 08-FEB-1996.  
PF 21-JUL-1995; U09234.  
PR 21-JUL-1994; US-278634.  
PR 23-NOV-1994; US-344824.  
PR 30-MAY-1995; US-452843.  
PA (CYTE-) CYTEL CORP.  
PI Sette A, Sidney J;  
DR WPI; 96-116784/12.  
PT Composn. comprising immunogenic peptide with supermotif allowing more  
than one HLA mol. to bind - used to induce CTL response in patient  
and for in vivo and ex vivo therapeutic and diagnostic applications  
PS Claim 2; Page 26; 32pp; English.  
CC The sequences given in R89362-82 are immunogenic peptides which were  
use in the composition of the invention. The composition comprises  
an immunogenic peptide of 9-10 residues with a supermotif which  
allows binding of more than one HLA molecule. It pref. comprises  
two conserved residues, a first at the 2nd position from the N-  
terminal is Pro, and a 2nd at the C-terminal is Met. These peptides  
are used to induce a CTL response in a patient. They are also  
useful in compositions for in vivo and ex vivo therapeutic and  
diagnostic applications, e.g the treatment of cancer and viral  
infections, e.g. hepatitis B and C.  
SQ Sequence 9 AA;

Query Match 100.0%; Score 73; DB 1; Length 9;  
Best Local Similarity 100.0%; Pred. No. 1.23e-01;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 1 IPYPIVRKL 9  
QY 1 IPYPIVRKL 9

RESULT 2  
ID R89370 standard; peptide; 9 AA.  
AC R89370;  
DT 18-SEP-1996 (first entry)

DE Cw6 consensus peptide derived immunogenic peptide #2.  
KW Immunogenic peptide; supermotif; HLA molecule; CTL response;  
KW therapeutic; diagnostic; cancer; viral infection; hepatitis B;  
KW hepatitis C.  
OS Synthetic.  
PN WO9603140-A1.  
PD 08-FEB-1996.  
PF 21-JUL-1995; U09234.  
PR 21-JUL-1994; US-278634.  
PR 23-NOV-1994; US-344824.  
PR 30-MAY-1995; US-452843.  
PA (CYTE-) CYTEL CORP.  
PI Sette A, Sidney J;  
DR WPI; 96-116784/12.  
PT Compsn. comprising immunogenic peptide with supermotif allowing more  
PT than one HLA mol. to bind - used to induce CTL response in patient  
PT and for in vivo and ex vivo therapeutic and diagnostic applications  
PS Claim 2; Page 26; 3pp; English.  
CC The sequences given in R89362-82 are immunogenic peptides which were  
CC use in the composition of the invention. The composition comprises  
CC an immunogenic peptide of 9-10 residues with a supermotif which  
CC allows binding of more than one HLA molecule. It pref. comprises  
CC two conserved residues, a first at the 2nd position from the N-  
CC terminal is Pro, and a 2nd at the C-terminal is Met. These peptides  
CC are used to induce a CTL response in a patient. They are also  
CC useful in compositions for in vivo and ex vivo therapeutic and  
CC diagnostic applications, e.g. the treatment of cancer and viral  
CC infections, e.g. hepatitis B and C.  
SQ Sequence 9 AA;

Query Match 89.0%; Score 65; DB 1; Length 9;  
Best Local Similarity 88.9%; Pred. No. 1.27e+00;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 1 IPIPIVRS 9  
QY 1 IPIPIVRKL 9

RESULT 3  
ID R89371 standard; peptide; 9 AA.

AC R89371;  
DT 18-SEP-1996 (first entry)  
DE Cw6 consensus peptide derived immunogenic peptide #3.  
KW Immunogenic peptide; supermotif; HLA molecule; CTL response;  
KW therapeutic; diagnostic; cancer; viral infection; hepatitis B;  
KW hepatitis C.  
OS Synthetic.  
PN WO9603140-A1.  
PD 08-FEB-1996.  
PF 21-JUL-1995; U09234.  
PR 21-JUL-1994; US-278634.  
PR 23-NOV-1994; US-344824.  
PR 30-MAY-1995; US-452843.  
PA (CYTE-) CYTEL CORP.  
PI Sette A, Sidney J;  
DR WPI; 96-116784/12.  
PT Compsn. comprising immunogenic peptide with supermotif allowing more  
PT than one HLA mol. to bind - used to induce CTL response in patient  
PT and for in vivo and ex vivo therapeutic and diagnostic applications  
PS Claim 2; Page 26; 3pp; English.  
CC The sequences given in R89362-82 are immunogenic peptides which were  
CC use in the composition of the invention. The composition comprises  
CC an immunogenic peptide of 9-10 residues with a supermotif which  
CC allows binding of more than one HLA molecule. It pref. comprises  
CC two conserved residues, a first at the 2nd position from the N-  
CC terminal is Pro, and a 2nd at the C-terminal is Met. These peptides  
CC are used to induce a CTL response in a patient. They are also  
CC useful in compositions for in vivo and ex vivo therapeutic and  
CC diagnostic applications, e.g. the treatment of cancer and viral  
CC infections, e.g. hepatitis B and C.  
SQ Sequence 9 AA;

Query Match 72.6%; Score 53; DB 1; Length 9;  
Best Local Similarity 85.7%; Pred. No. 3.66e+01;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 1 IPEPIVR 7  
QY 1 IPEPIVR 7

RESULT 4

ID W8511 standard; Protein; 133 AA.  
AC W8511;  
DT 29-DEC-1997 (first entry)  
DE Product of clone LI05.  
KW J5; J422; LI05; H174-10; H174-43; B18; cytokine; PBMC;  
KW peripheral blood mononuclear cell; disintegrin; metallo-protein;  
KW Drosophila; leucine-rich repeat; monocyte; chemoattractant;  
KW IP-10; CRG-2; CTLA-8; herpesvirus; Salmir.  
OS Mus musculus.  
PN W09707198-A2.  
PD 27-FEB-1997.  
PF 08-AUG-1996; U12897.  
PR 08-AUG-1996; WO-U12897.  
PA (GEMV) GENETICS INST INC.  
PI Carlin M, Jacobs K, Kelleher K, McCoy JM;  
DR WPI; 97-165283/15.  
DR N-PSDB; T87429.  
PT Polynucleotide(s) encoding proteins for treating, preventing and  
PT ameliorating medical conditions - obtained from human activated  
PT peripheral blood mononuclear cell, and murine adult thymus libraries  
PS Claim 21; Page 44-45; 61pp; English.  
CC This sequence was isolated from a murine adult thymus library using  
CC a trap selecting for nucleotides encoding secreted proteins, and  
CC encodes a protein having homology to various monocyte and other  
CC chemoattractant proteins.  
SQ Sequence 133 AA;

Query Match 69.9%; Score 51; DB 1; Length 133;  
Best Local Similarity 85.7%; Pred. No. 6.26e+01;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 40 IPIPIVR 46  
QY 1 IPIPIVR 7

RESULT 5

ID W50884 standard; Protein; 133 AA.  
AC W50884;  
DT 09-SEP-1998 (first entry)  
DE Amino acid sequence of mouse 6CKine protein.  
KW Mouse; 6CKine gene; m6CKine; chemokine; mpf4; MCTAP3;  
KW h6CKine; Chrl9Kine; cancer; degenerative condition; antibody;  
KW immuno assay; forensic assay; in situ assay.  
OS Mus sp.  
FH Key Location/Qualifiers  
FT Peptide 1..23 /note= "signal peptide"  
FT Protein 24..133 /note= "mature protein"  
FN W09814581-A1.  
PD 09-APR-1998.  
PF 02-OCT-1997; U17122.  
PR 28-AUG-1997; US-058007.  
PR 02-OCT-1996; US-027242.  
PR 09-OCT-1996; US-028042.  
PA (SCHE) SCHERING CORP.  
PI Hedrick JA, Zlotnik A;  
DR WPI; 98-240086/21.  
DR N-PSDB; V07113.  
PT Mouse and human CC and CXK chemokine(s) - useful to modulate  
PT physiology or development of cells to treat, e.g. cancerous or  
PT degenerative conditions

PS Claim 1; Pages 78-79; 89pp; English.  
 CC This is the amino acid sequence of the mouse 6CKine (m6CKine) gene, a  
 CC chemokine. It is used in the method of the invention where mouse and  
 CC human CC and CXC chemokines, designated mpf4, mCTAP3, m6CKine, h6CKine  
 CC and Chrl9kine are used to modulate the physiology or the development  
 CC of cells to treat, cancerous or degenerative conditions. The  
 CC chemokines can also be used to generate antibodies, useful in  
 CC immunoassays to measure chemokines, while the nucleic acid sequences  
 CC may be used as components in forensic assays or in situ assays to  
 CC detect chromosomal abnormalities.  
 SQ Sequence 133 AA;

Query Match 69.9%; Score 51; DB 1; Length 133;  
 Best Local Similarity 85.7%; Pred. No. 6.26e+01;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 40 IPYSIVR 46  
 :|||:|  
 QY 1 IPYPIVR 7

## RESULT 6

ID P70388 standard; protein; 356 AA.  
 AC P70388;  
 DT 14-JAN-1991 (first entry)  
 DE D-amino acid oxidase.  
 KW D-amino acid oxidase; Trigonopsis variabilis; cephalosporin;  
 KW oxidative deamination.  
 OS Trigonopsis variabilis.  
 PN J62262394-A.  
 PD 16-NOV-1987.  
 PF 12-MAY-1986; JP-106663.  
 PR 12-MAY-1986; JP-106663.  
 PA (ASAH) ASAH CHEMICAL IND KK.  
 DR WPI; 87-359677/51.  
 DR N-PSDB; N70609.  
 PT DNA fragment encoding D-amino acid oxidase - which is a useful  
 PT enzyme for the catalytic oxidative deamination of D-amino acids.  
 PS Claim 1; page 583-4; 12pp; Japanese.  
 CC D-amino acid oxidase catalyses the oxidative deamination of D-amino  
 CC acids. It is used in the sepn. of L-amino acids from racemates,  
 CC in the prepn. of ketoic acid from D-amino acid, in amino acid  
 CC analysis, etc. The enzyme can oxidise cephalosporin C to  
 CC 7-beta-(5-carboxy-5-oxopentanamide)cephalosporanic acid, which  
 CC reacts with hydrogen peroxide to give 7-beta-(4-carboxybutanamide)-  
 CC cephalosporanic acid. These cpds. are important intermediates for  
 CC synthesis of cephalosporin type antibiotics.  
 SQ Sequence 356 AA;

Query Match 68.5%; Score 50; DB 1; Length 356;  
 Best Local Similarity 55.6%; Pred. No. 8.17e+01;  
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 66 VSYPIREL 74  
 :|||:|  
 QY 1 IPYPIVR 9

## RESULT 7

ID R04066 standard; Protein; 356 AA.  
 AC R04066;  
 DT 03-SEP-1990 (first entry)  
 DE T. variabilis D-amino acid oxidase gene product.  
 KW D-amino acid oxidase; cephalosporin; cephem; E.coli.  
 OS Trigonopsis variabilis.  
 PN EP-364275-A.  
 PD 18-APR-1990.  
 PF 12-OCT-1989; 310483.  
 PR 13-OCT-1988; JP-260332.  
 PA (FUJI) Fufisawa Pharm KK.  
 PI Isogai T, Ono H, Kojo H;  
 DR WPI; 90-117771/16.  
 PT D-amino acid oxidase, prodn. -

PT by culture of E.coli transformants contg. expression vectors  
 PT originated from Fusarium solani M-0718.  
 PS Disclosure; Fig 9; 38pp; English.  
 CC E.coli transformed to express DAO, which catalyses the enzymatic  
 CC conversion of cephalosporin C to 7-beta-(5-carboxy-5-  
 CC oxopentanamide)cephalosporanic acid (keto-7ACA). 7ACA is an  
 CC important starting point for the production of cephem  
 CC antibiotics.  
 SQ Sequence 356 AA;

Query Match 68.5%; Score 50; DB 1; Length 356;  
 Best Local Similarity 55.6%; Pred. No. 8.17e+01;  
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 65 VSYPIREL 73  
 :|||:|  
 QY 1 IPYPIVR 9

## RESULT 8

ID W47206 standard; Protein; 1199 AA.  
 AC W47206;  
 DT 21-JUL-1998 (first entry)  
 DE Bos taurus tubulin-folding cofactor D.  
 KW Tubulin folding; cofactor; alpha-tubulin; beta-tubulin; unfolded;  
 KW folded; treatment; hyper-proliferative diseases; cancer; gout.  
 OS Bos taurus.  
 PN W09804587-A1.  
 PD 05-FEB-1998.  
 PF 25-JUL-1997; U14076.  
 PR 25-JUL-1996; US-023089.  
 PA (UYNY) UNIV NEW YORK STATE.  
 PI Cowan NJ;  
 DR WPI; 98-130618/12.  
 DR N-PSDB; V17086.  
 PT New isolated cofactor(s) for tubulin folding - are useful as targets  
 PT for identifying agents which interfere with folding in the treatment  
 PT of hyper-proliferative diseases such as cancer  
 PS Claim 3; Pages 48-52; 87pp; English.  
 CC The sequence is that of bovine tubulin-folding cofactor D.  
 CC It may be useful as a target for interfering with the  
 CC production of productively folded alpha- and beta-tubulins.  
 CC Since tubulin function is essential for cell division and  
 CC proliferation, agents which interfere with tubulin function  
 CC can serve as useful antiproliferative compounds. Such interfering  
 CC agents have potential utility as agents for the treatment of  
 CC hyperproliferative diseases such as cancer and the treatment  
 CC of gout.  
 SQ Sequence 1199 AA;

Query Match 67.1%; Score 49; DB 1; Length 1199;  
 Best Local Similarity 57.1%; Pred. No. 1.06e+02;  
 Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 1128 PFPVIRK 1134  
 :|||:|  
 QY 2 PYPIVRK 8

## RESULT 9

ID R77604 standard; Protein; 1676 AA.  
 AC R77604;  
 DT 15-MAR-1996 (first entry)  
 DE Pro-C5 polypeptide  
 KW Complement C5; haemolysis; kidney; glomerulonephritis;  
 KW monoclonal antibody; antiinflammatory; antibody engineering;  
 KW humanised antibody.  
 OS Homo sapiens.  
 PH Key Location/Qualifiers  
 FT peptide 1..18  
 FT /label= Sig\_peptide  
 FT protein 19..673  
 FT /label= Beta-chain

```
FT cleavage_site 673..674
FT cleavage_site 677..678
FT peptide 674..677
FT label= Cleavage_peptide
FT protein 678..1676
FT /label= Alpha-chain
FT /note= "amino acids 872-892 (854-874 of
FT the mature protein) comprise the KSSKS
FT epitope"
FT peptide 678..751
FT /label= C5a
FT cleavage_site 751..752
FT /label= Convertase_cleavage_site
FT modified_site 911
FT /label= N-glycosylation_site
FT modified_site 1115
FT /label= N-glycosylation_site
FT modified_site 1630
FT /label= N-glycosylation_site
FT W09529697-A1.
FT PN 09-NOV-1995.
FT PD 01-MAY-1995; U05688.
FT PR 02-MAY-1994; US-236208.
FT PA (ALEX-) ALEXION PHARM INC.
FT PI Evans MJ, Mattis L, Mueller EE, Nye SH, Rollins S;
FT PI Rother RP, Springhorn J P, Squinto SP, Thomas TC;
FT PI Wang Y, Wilkins JA;
FT DR WPI; 95-392323/50.
FT PT Treating glomerulonephritis with antibody against complement C5
FT component - to inhibit complement induced cell lysis
FT Example 13; Page 82-92; 181pp; English.
FT PS The cDNA sequence of the complement C5 gene transcript predicts a
FT CC secreted pro-C5 precursor of 1676 amino acids (R77604). C5 is a
FT CC beta-globulin heterodimer thought to play a role in the pathogenesis
FT CC of glomerulonephritis (GN). Cleavage of the C5 alpha-chain
FT CC by a convertase enzyme generates anaphylatoxic C5a. Monoclonal
FT CC and humanised recombinant antibodies that recognise the alpha-chain
FT CC KSSKC epitope (R77605) block C5a generation, thereby reducing
FT CC glomerular inflammation and kidney dysfunction associated with GN.
FT SQ Sequence 1676 AA;

Query Match 67.1%; Score 49; DB 1; Length 1676;
Best Local Similarity 71.4%; Pred. No. 1.06e+02;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 829 IPYSWR 835
QY 1 IPYDVR 7

RESULT 10
ID W44125 standard; peptide; 159 AA.
AC W44125;
DE 28-APR-1998 (first entry)
DE Streptococcus pneumoniae leucyl tRNA synthetase fragment.
KW Leucyl tRNA synthetase; leucyl polypeptide; vaccine;
KW genetic immunisation; antibacterial; antibiotic; otitis media;
KW conjunctivitis; pneumonia; bacteraemia; meningitis; sinusitis;
KW pleural empysema; endocarditis; gene therapy.
OS Streptococcus pneumoniae.
FT Key Location/Qualifiers
FT Misc_difference 53
FT /label= Unspecified
FT /note= "encoded by NAA"
FT PN W09739022-A1.
FT PD 23-OCT-1997.
FT PF 18-APR-1997; U06875.
FT PR 18-APR-1996; GB-007993.
FT PA (SWIK ) SMITHKLINE BEECHAM CORP.
FT PA (SWIK ) SMITHKLINE BEECHAM PLC.
FT PI Lawlor EJ;
FT DR WPI; 97-526396/48.
FT DR N-PSDB; V12059.

PT Streptococcus pneumoniae leucyl tRNA synthetase - useful to produce
PT antibodies or to screen for (antagonists with antibacterial
PT activity, e.g. to diagnose and treat meningitis, pneumonia, etc.
PT Claim 12; Page 39-40; 48pp; English.
PS The present sequence represents a leucyl tRNA synthetase (leuS) fragment
CC from Streptococcus pneumoniae. The leuS polypeptides, antagonists,
CC antibodies and related nucleic acids can be used for diagnosis and
CC treatment of bacterial diseases. In particular, they are directed
CC towards Streptococcus pneumoniae infections causing otitis media,
CC conjunctivitis, pneumonia, bacteraemia, meningitis, sinusitis, pleural
CC empysema and endocarditis. LeuS polypeptides, or vectors for their
CC expression, can be used prophylactically in vaccines to raise an
CC antibody and/or T cell immune response against these same diseases.
CC Additionally, the new polypeptides allow agonists and antagonists of
CC leuS to be identified using standard binding assays. The compounds which
CC are identified may have useful bacteriostatic and/or bacteriocidal
CC activity. 159 AA;
SQ Sequence 159 AA;

Query Match 65.8%; Score 48; DB 1; Length 159;
Best Local Similarity 83.3%; Pred. No. 1.38e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 136 YPVVRK 141
QY 3 YPIVRK 8

RESULT 11
ID W25084 standard; Protein; 311 AA.
AC W25084;
DE 30-DEC-1997 (first entry)
DE Haemophilus influenzae htrB polypeptide.
KW Vaccine; htrB gene; Gram-negative bacterium; non-toxic mutant;
KW pathogen; endotoxin; diagnosis; passive immunisation.
OS Haemophilus influenzae strain 2019.
PN W09719688-A1.
PD 05-JUN-1997.
PF 27-NOV-1996; U18984.
PR 01-DEC-1995; US-565943.
PA (AMCY ) AMERICAN CYANAMID CO.
PA (REGC ) UNIV CALIFORNIA.
PA (IOWA ) UNIV IOWA RES FOUND.
PI Apicella MA, Arumugham R, Gibson BW, Lee N, Sunshine MG;
DR WPI; 97-310355/28.
DR N-PSDB; T79708.
PT New Gram-negative bacterial pathogen vaccines - comprising a htrB
PT mutant or an endotoxin isolated from an htrB mutant optionally
PT conjugated to a carrier protein.
PS Example 1; Page 61-62; 79pp; English.
CC This polypeptide comprises the htrB gene product (see also T79708)
CC of Haemophilus influenzae strain 2019. A claimed vaccine
CC formulation contains as an active ingredient an htrB mutant of a
CC Gram-negative bacterial pathogen (GNBP), endotoxin isolated from an
CC htrB mutant (A) of a GNBP, endotoxin isolated from (A) conjugated
CC to a carrier protein, or (A) which has been genetically engineered
CC to express at least one heterologous vaccine antigen, where (A)
CC lacks one or more secondary acyl chains of lipid A contained in the
CC GNBP resulting in reduced toxicity when compared to lipid A of the
CC GNBP. Also claimed is a method for producing endotoxin-specific
CC antisera for diagnostic assays, or for passive immunisation,
CC comprising immunising an individual with a vaccine formulation
CC comprising an active ingredient as above, and collecting antibodies
CC produced from the immunised individual.
SQ Sequence 311 AA;

Query Match 65.8%; Score 48; DB 1; Length 311;
Best Local Similarity 55.6%; Pred. No. 1.38e+02;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 34 LPYPIRHI 42
QY 1 IPYPIVRKL 9
```



```

RESULT 12
ID W32114 standard; Protein; 833 AA.
AC W32114;
DT 28-APR-1998 (first entry)
DE Streptococcus pneumoniae leucyl tRNA synthetase.
KW Leucyl tRNA synthetase; leucyl polypeptide; vaccine;
KW genetic immunisation; antibacterial; antibiotic; otitis media;
KW conjunctivitis; pneumonia; bacteraemia; meningitis; sinusitis;
KW pleural emphysema; endocarditis; gene therapy.
OS Streptococcus pneumoniae.
PN W09739022-AL.
PD 23-OCT-1997.
PR 18-APR-1997; U06875.
PR 18-APR-1996; GB-007993.
PA (SMIK ) SMITHKLINE BEECHAM CORP.
PA (SMIK ) SMITHKLINE BEECHAM PLC.
PI Lawlor EJ;
PI WPI; 97-526396/48.
DR N-PSDB; T88991.
DR Streptococcus pneumoniae leucyl tRNA synthetase - useful to produce
PT antibodies or to screen for (ant)agonists with antibacterial
PT activity, e.g. to diagnose and treat meningitis, pneumonia, etc.
PS Claim 12; Page 36-38; 48pp; English.
CC Streptococcus pneumoniae. The leucyl polypeptides, antagonists, antibodies
CC and related nucleic acids can be used for diagnosis and treatment of
CC bacterial diseases. In particular, they are directed towards
CC Streptococcus pneumoniae infections causing otitis media,
CC conjunctivitis, pneumonia, bacteraemia, meningitis, sinusitis, pleural
CC emphysema and endocarditis. Leucyl polypeptides, or vectors for their
CC expression, can be used prophylactically in vaccines to raise an
CC antibody and/or T cell immune response against these same diseases.
CC Additionally, the new polypeptides allow agonists and antagonists of
CC leucyl to be identified using standard binding assays. The compounds which
CC are identified may have useful bacteriostatic and/or bacteriocidal
CC activity.
SQ Sequence 833 AA;

Query Match 65.88; Score 48; DB 1; Length 833;
Best Local Similarity 83.38; Pred. No. 1.38e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 182 YPVVRK 187
|||
QY 3 YPIVRK 8

RESULT 13
ID W73419 standard; Protein; 216 AA.
AC W73419;
DT 19-FEB-1999 (first entry)
DE Human secreted protein encoded by Gene No. 23.
KW Secreted protein; human; protein therapy; gene therapy; blood disorder;
KW pathological condition; diagnosis; cancer; neurological disorder;
KW developmental abnormality; foetal deficiency; leukaemia; hepatic disease;
KW immune system disorder; Alzheimer's disease; cognitive disorder;
KW schizophrenia; prostate disease; autoimmune disorder; AIDS.
OS Homo sapiens.
FH Key Location/Qualifiers
FT MISC_difference 216 /note= "unspecified amino acid"
FT W09854206-AL.
PD 03-DEC-1998.
PF 28-MAY-1998; U10868.
PR 29-AUG-1997; US-056296.
PR 30-MAY-1997; US-044039.
PR 30-MAY-1997; US-048093.
PR 30-MAY-1997; US-048101.
PR 30-MAY-1997; US-048190.
PR 30-MAY-1997; US-048356.
PR 30-MAY-1997; US-050935.

RESULT 14
ID W53896 standard; Protein; 319 AA.
AC W53896;
DT 28-AUG-1998 (first entry)
DE Human G-protein coupled receptor HLYA261.
KW HLYA261; G-protein coupled receptor; human; therapy;
KW diagnosis; infection; HIV-1; HIV-2; pain; cancer; anorexia;
KW bulimia; asthma; Parkinson's disease; acute heart failure;
KW atherosclerosis; hypotension; hypertension; urinary retention;
KW osteoporosis; angina pectoris; myocardial infarction; ulcer;
KW allergy; benign prostatic hypertrophy; neurological disorder;
KW psychosis; anxiety; schizophrenia; manic depression; delirium;
KW dementia; mental retardation; dyskinesia; Huntington's disease;
KW Gilles de la Tourette's syndrome.
OS Homo sapiens.
PN EP-837128-A2.
PD 22-APR-1998.
PF 16-OCT-1997; 308207.
PR 21-OCT-1996; US-734349.
PA (SMIK ) SMITHKLINE BEECHAM CORP.
PI Bergsma DJ, Ellis CE;
PI WPI; 98-219111/20.
DR N-PSDB; V23658.
DR DNA encoding G-protein coupled receptor protein - useful for
PT producing recombinant peptides and in gene therapy
PS Claim 13; Fig 1a-c; 38pp; English.
CC This polypeptide comprises HLYA261, a novel human G-protein coupled
CC receptor containing 7 hydrophobic regions that may represent
CC a membrane spanning domain. Its amino acid sequence was deduced from
CC a cDNA clone (see V23658) isolated from a human leukocyte cDNA
CC library. This polynucleotide can be utilised in the recombinant
CC production of HLYA261 in host cells. HLYA261 polypeptides may be
CC employed for therapeutic purposes, including treatment of bacterial,
CC fungal, protozoan and viral infections, particularly infections

Query Match 64.48; Score 47; DB 1; Length 216;
Best Local Similarity 71.48; Pred. No. 1.80e+02;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 140 VPXHIVR 146
|||
QY 1 IPIPIVR 7

RESULT 15
ID W53896 standard; Protein; 319 AA.
AC W53896;
DT 28-AUG-1998 (first entry)
DE Human G-protein coupled receptor HLYA261.
KW HLYA261; G-protein coupled receptor; human; therapy;
KW diagnosis; infection; HIV-1; HIV-2; pain; cancer; anorexia;
KW bulimia; asthma; Parkinson's disease; acute heart failure;
KW atherosclerosis; hypotension; hypertension; urinary retention;
KW osteoporosis; angina pectoris; myocardial infarction; ulcer;
KW allergy; benign prostatic hypertrophy; neurological disorder;
KW psychosis; anxiety; schizophrenia; manic depression; delirium;
KW dementia; mental retardation; dyskinesia; Huntington's disease;
KW Gilles de la Tourette's syndrome.
OS Homo sapiens.
PN EP-837128-A2.
PD 22-APR-1998.
PF 16-OCT-1997; 308207.
PR 21-OCT-1996; US-734349.
PA (SMIK ) SMITHKLINE BEECHAM CORP.
PI Bergsma DJ, Ellis CE;
PI WPI; 98-219111/20.
DR N-PSDB; V23658.
DR DNA encoding G-protein coupled receptor protein - useful for
PT producing recombinant peptides and in gene therapy
PS Claim 13; Fig 1a-c; 38pp; English.
CC This polypeptide comprises HLYA261, a novel human G-protein coupled
CC receptor containing 7 hydrophobic regions that may represent
CC a membrane spanning domain. Its amino acid sequence was deduced from
CC a cDNA clone (see V23658) isolated from a human leukocyte cDNA
CC library. This polynucleotide can be utilised in the recombinant
CC production of HLYA261 in host cells. HLYA261 polypeptides may be
CC employed for therapeutic purposes, including treatment of bacterial,
CC fungal, protozoan and viral infections, particularly infections
```

Job time : 40 secs.

CC caused by HIV-1 and HIV-2, pain, cancers, anorexia, bulimia,  
CC asthma, Parkinson's disease, acute heart failure, atherosclerosis,  
CC hypertension, hypotension, urinary retention, osteoporosis, angina  
CC pectoris, myocardial infarction, ulcers, allergies, benign  
CC prostatic hypertrophy and psychotic and neurological disorders  
CC including anxiety, schizophrenia, manic depression, delirium,  
CC dementia or severe mental retardation, and dyskinesias, such as  
CC Huntington's disease or Gilles de la Tourette's syndrome. The  
CC polypeptide can also be used in a claimed method for identifying  
CC compounds which bind to and activate or inhibit a receptor for  
CC H1RAZ61. Also disclosed are diagnostic assays for detecting  
CC diseases related to altered concentrations of H1RAZ61 polypeptides.  
SQ Sequence 319 AA;

Query Match 64.4%; Score 47; DB 1; Length 319;  
Best Local Similarity 71.4%; Pred. No. 1.80e+02;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

DB 237 VPYHIVR 243  
: || |||  
QY 1 IPYPIVR 7

RESULT 15  
ID W69735 standard; Protein; 319 AA.  
AC W69735;  
DT 26-OCT-1998 (first entry)  
DE Human C5a-like protein.  
KW Human; C5a-like protein; HCOR; diagnosis; complement activation;  
KW inflammation; immunodeficiency; brain de-myelination; neurodegeneration;  
KW allergic reaction; asthma; adult respiratory distress syndrome;  
KW autoimmune disorder; rheumatoid arthritis; systemic lupus erythematosus;  
KW glomerulonephritis; Crohn's disease; cancer; haemodialysis.  
OS Homo sapiens.  
PN W09833908-A1.  
PD 06-AUG-1998.  
PF 20-JAN-1998; U01182.  
PR 31-JAN-1997; US-791974.  
PA (INCY-) INCYTE PHARM INC.  
PI Bandman O, Coleman R;  
DR WPI: 98-437462/37.  
DR N-PSDB; V50491.  
PT Isolated human C5a-like receptor - used to develop products for  
PT diagnosis, prevention and treatment of disorders associated with  
PT complement activation, particularly inflammation  
PS Claim 1; Page 42-43; 59pp; English.  
CC The present sequence represents human C5a-like protein (HCOR). The HCOR  
CC has similarity to human C5a receptor. Products from the present invention  
CC can be used for the diagnosis, prevention, or treatment of diseases  
CC associated with complement activation. The HCOR and agonists can be used  
CC to induce an inflammatory response in a subject who has a diminished  
CC inflammatory response as a result of conditions such as complement  
CC deficiency, immunodeficiency and impaired wound healing. Antagonists or  
CC inhibitors of HCOR can be used to prevent inflammation in, e.g. brain  
CC de-myelination and neurodegeneration, allergic reactions, asthma and  
CC adult respiratory distress syndrome, autoimmune disorders such as  
CC rheumatoid arthritis, systemic lupus erythematosus, glomerulonephritis,  
CC and Crohn's disease, post ischaemic myocardial inflammation and necrosis,  
CC skin diseases, septic shock, and inflammatory complications of cancer,  
CC haemodialysis and extracorporeal circulation, infection and trauma. The  
CC products can also be used for detection and drug screening.  
SQ Sequence 319 AA;

Query Match 64.4%; Score 47; DB 1; Length 319;  
Best Local Similarity 71.4%; Pred. No. 1.80e+02;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

DB 237 VPYHIVR 243  
: || |||  
QY 1 IPYPIVR 7

Search completed: Fri Apr 14 23:43:00 2000

\*\*\*\*\*  
 W P S R L  
 (TM)  
 \*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
 Copyright (c) 1993-1998 University of Edinburgh, U.K.  
 Distribution rights by Oxford Molecular Ltd

MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm  
 Run on: Fri Apr 14 23:43:18 2000; MasPar time 3.34 Seconds  
 Tabular output not generated. 107.808 Million cell updates/sec

Title: >US-08-452-843-8  
 Description: (1-9) from US08452843.pap  
 Perfect Score: 73  
 Sequence: 1 IYPPIVRKL 9

Scoring table: PAM 150  
 Gap 15

Searched: 122810 seqs, 40068593 residues

Post-processing: Minimum Match 0%  
 Listing first 45 summaries

Database: p1r62  
 1.p1r1 2.p1r2 3.p1r3 4.p1r4

Statistics: Mean 24.378; Variance 31.755; scale 0.768

Pred. NO. is the number of results predicted by chance to have a  
 score greater than or equal to the score of the result being printed,  
 and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	59	80.8	410	2	D70884	3.28e-01
2	55	75.3	1679	2	S48385	hypothetical protein
3	53	72.6	52	2	F70083	hypothetical protein
4	53	72.6	209	2	S3179	transforming protein
5	53	72.6	487	1	Q8ECS	YJJE protein - Escher
6	52	71.2	451	2	S75569	hypothetical protein
7	52	71.2	880	2	S35926	DNA-directed RNA poly
8	51	69.9	75	2	H69915	hypothetical protein
9	51	69.9	197	2	C64422	hypothetical protein
10	51	69.9	340	2	S62493	hypothetical protein
11	51	69.9	520	2	F70350	recombination protein
12	51	69.9	607	1	NUOTB	glucose-6-phosphate i
13	51	69.9	766	2	A56394	pyocin S3 - pseudomon
14	50	68.5	357	2	G70577	probable dihydrorota
15	50	68.5	615	2	H64769	preprotein translocas
16	49	67.1	125	2	S23541	hypothetical protein
17	49	67.1	183	2	E54506	adenine phosphoribosy
18	49	67.1	227	2	C69432	hypothetical protein
19	49	67.1	401	2	G71018	hypothetical protein
20	49	67.1	1061	1	S27311	ribonuclease E (EC 3.
21	49	67.1	1676	1	CSHU	complement C5 precurs
22	49	67.1	1680	1	CSMS	complement C5 precurs
23	48	65.8	100	2	S44892	ZK112.4 protein - Cae

48 65.8 156 2 C64486  
 48 65.8 219 2 S75541  
 25 48 65.8 305 2 S35991  
 26 48 65.8 315 2 D64127  
 27 48 65.8 325 2 S26216  
 28 48 65.8 337 2 T02532  
 29 48 65.8 377 2 S25156  
 30 48 65.8 407 2 B32306  
 31 48 65.8 420 2 A55145  
 32 48 65.8 540 2 A55145  
 33 48 65.8 617 2 T56530  
 34 48 65.8 617 2 R39748  
 35 48 65.8 663 2 T03217  
 36 48 65.8 762 1 NNNC2  
 37 48 65.8 1381 2 S55619  
 38 47 64.4 196 2 E69042  
 39 47 64.4 238 2 D71189  
 40 47 64.4 268 2 S39711  
 41 47 64.4 329 1 AJZJ02  
 42 47 64.4 342 2 G70712  
 43 47 64.4 364 2 S77360  
 44 47 64.4 665 2 T02793  
 45 47 64.4 682 1 HHBYK2

## ALIGNMENTS

RESULT 1

ENTRY  
 TITLE  
 ORGANISM  
 DATE  
 REFERENCES  
 #authors  
 D70884 #type complete  
 Probable Acyl-CoA Dehydrogenas - Mycobacterium tuberculosis  
 (strain H37RV)  
 #formal\_name Mycobacterium tuberculosis  
 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change  
 17-Jul-1998  
 D70884  
 A70500  
 Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.V.; Eigmeier, K.; Gas, S.; Barry III, C.E.; Tekle, F.; Badcock, K.; Basham, D.; Brown, D.; Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Hornsby, T.; Jagels, K.; Krogh, A.; McLean, J.; Moule, S.; Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Squires, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
 #journal Nature (1998) 393:537-544  
 #title Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence.  
 #cross-references MUID:98295987  
 #accession D70884  
 #status preliminary: nucleic acid sequence not shown;  
 translation not shown  
 ##molecule\_type DNA  
 ##residues 1-410 #label COL  
 ##cross-references GB:AL008967; GB:AL123456; NID:g3261491; PID:ell73919;  
 PID:g2624311  
 ##experimental\_source strain H37RV  
 GENETICS  
 #gene fadE21  
 SUMMARY  
 #length 410 #molecular-weight 44743 #checksum 7911  
 Query Match 80.8%; Score 59; DB 2; Length 410;  
 Best Local Similarity 87.5%; Pred. No. 3.28e-01;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 Db 37 PYPIARXL 44  
 |||||  
 Qy 2 PYPIVRKL 9  
 RESULT 2  
 ENTRY  
 TITLE  
 S48385 #type complete  
 hypothetical protein Y1149c - yeast (Saccharomyces

```
cerevisiae)
#formal_name Saccharomyces cerevisiae
#sequence_revision 02-Dec-1994 #text_change
02-Dec-1994 #sequence_revision 02-Dec-1994 #text_change
DATE 21-Nov-1997
ACCESSIONS S48385
REFERENCE S48310
#authors Churcher, C.
#submission submitted to the EMBL Data Library, September 1994
#accession S48385
#molecule_type DNA
#residues 1-1679 ##label CHU
##cross-references GB:247047; EMBL:Z38059; NID:G603997; PID:G763197;
MIPS:Y1149C
GENETICS
#map_position 9L
#length 1679 #molecular-weight 195140 #checksum 5165
SUMMARY
Query Match 75.3%; Score 55; DB 2; Length 1679;
Best Local Similarity 55.6%; Pred. No. 2.13e+00;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
Db 19 VTYPIVRLK 27
:|||||
QY 1 IPYPIVRLK 9

RESULT 3
ENTRY F70083
TITLE #type complete
ORGANISM hypothetical protein yxzF - Bacillus subtilis
DATE 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change
F70083
ACCESSIONS A69580
REFERENCE Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.;
Alloni, G.; Azevedo, V.; Bertero, M.G.; Bessieres, P.;
Bolotin, A.; Borchert, S.; Boriss, R.; Boursier, L.; Brans,
A.; Braun, M.; Brignell, S.C.; Bron, S.; Brouillet, S.;
Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
Choi, S.K.; Codani, J.J.; Connerton, I.F.; Cummings, N.J.;
Daniel, R.A.; Denizot, F.; Devine, K.M.; Duesterhoeft, A.;
Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Erington, J.;
Fabret, C.; Ferrari, E.; Foulger, D.; Fritz, C.; Fujita,
M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galleron, N.; Ghim,
S.Y.; Glaser, P.; Goffeau, A.; Gollightly, E.J.; Grandi, G.;
Guiseppi, G.; Guy, B.J.; Haga, K.; Haiech, J.; Harwood,
C.R.; Henaut, A.; Hilbert, R.; Hoisappel, S.; Hosono, S.;
Hullo, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karamata, D.;
Kasahara, Y.; Klaerr-Blanchard, M.; Klein, C.; Kobayashi,
Y.; Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.;
Kurita, K.; Lapidus, A.; Lardinois, S.; Lauber, J.;
Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.;
Mauel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno,
M.; Moestl, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly,
M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro,
V.; Pohl, T.M.; Portetle, D.; Porwolik, S.; Prescott, G.;
A.M.; Presecan, E.; Pujic, P.; Purnelle, B.; Rapoport, G.;
Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.;
Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon, E.;
Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.;
Sekowska, A.; Seror, S.J.; Serror, P.; Shin, B.S.; Soldo,
B.; Sorokin, A.; Taccioni, E.; Takagi, T.; Takahashi, H.;
Takemaru, K.; Takeuchi, M.; Tamakoshi, A.; Tanaka, T.;
Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.;
Vandenbol, M.; Vannier, F.; Vassarotti, A.; Viari, A.;
Wambutt, R.; Wedler, E.; Wedler, H.; Weitzenecker, T.;
Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto,
K.; Yata, K.; Yoshida, K.; Yoshikawa, H.F.; Zumstein, E.;
Yoshikawa, H.; Danchin, A.
Nature (1997) 390:249-256
The complete genome sequence of the Gram-positive bacterium
Bacillus subtilis.
#cross-references MUID:98044033

#accession F70083
#status preliminary; nucleic acid sequence not shown;
translation not shown
#molecule_type DNA
#residues 1-52 ##label KUN
##cross-references GB:299123; GB:AL009126; NID:G2636240; PID:el186360;
PID:G2636396
##experimental_source strain 168
GENETICS
#gene yxzF
#length 52 #molecular-weight 5915 #checksum 5675
SUMMARY
Query Match 72.6%; Score 53; DB 2; Length 52;
Best Local Similarity 66.7%; Pred. No. 5.25e+00;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Db 25 VTYPIVRLK 33
:|||||
QY 1 IPYPIVRLK 9

RESULT 4
ENTRY S13179
TITLE #type complete
ORGANISM transforming protein (ras) - Geodia cydonium
DATE 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change
19-Dec-1998
S13179
ACCESSIONS S13179
REFERENCE Robitzki, A.; Schroeder, H.C.; Ugarkovic, D.; Kuchino, Y.;
Kurelec, B.; Gamulin, V.; Mueller, W.E.G.
Eur. J. Biochem. (1990) 192:499-506
#journal Regulated expression and phosphorylation of the 23-26-kDa ras
#title protein in the sponge Geodia cydonium.
#cross-references MUID:91006138
#accession S13179
#status preliminary
#molecule_type mRNA
#residues 1-209 ##label ROB
#note based on the evidence for Gln-trNA, the authors
translated the codon TAG as Gln; the sequence shown
follows the authors' translation
CLASSIFICATION #superfamily ras transforming protein; translation
factor Tu homology
KEYWORDS GTP binding; P-loop
FEATURE
10-17 #region nucleotide-binding motif A (P-loop)\
140-143 #region GTP-binding NKXD motif\
168-170 #region GTP-binding SAK/L motif\
16,17,58,140,141,
143,168 #binding_site Mg-GTP (Lys, Ser, Thr, Asn, Lys, Asp, Ser)
#status predicted
SUMMARY #length 209 #molecular-weight 23854 #checksum 3860
Query Match 72.6%; Score 53; DB 2; Length 209;
Best Local Similarity 66.7%; Pred. No. 5.25e+00;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Db 178 IPYSLVREL 186
:|||||
QY 1 IPYPIVRLK 9

RESULT 5
ENTRY Q0ECS
TITLE #type complete
ORGANISM yjyE protein - Escherichia coli
DATE 30-Jun-1988 #sequence_revision 31-Oct-1997 #text_change
17-Jul-1998
E65094; C29049
ACCESSIONS E65094; C29049
REFERENCE Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.;
Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.;
```

```

Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.;
Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao,
Y.
#journal Science (1997) 277:1453-1462
#title The complete genome sequence of Escherichia coli K-12.
#cross-references MUID:97428617
#accession E65094
#status nucleic acid sequence not shown; translation not shown
#molecule_type DNA
#residues 1-487 #label BLAT
#cross-references GB:AE000388; GB:U00096; NID:gl789441; PID:gl789444;
#experimental_source strain K-12, substrain MG1655
#reference A91573
#authors Nesin, M.; Lupski, J.R.; Svec, P.; Godson, G.N.
#journal Gene (1987) 51:149-161
#title Possible new genes as revealed by molecular analysis of a
5-kb Escherichia coli chromosomal region 5' to the
rpsU-dnaG-rpoD macromolecular-synthesis operon.
#cross-references MUID:87248073
#accession C29049
#molecule_type DNA
#residues 279-403, 'P', 405-411, 'RWRCKSRCRCSA' #label NES
GENETICS
#gene ygjE
#map_position 67 min
CLASSIFICATION #superfamily 2-oxoglutarate/malate translocator
transmembrane protein
FEATURE
11-27 #domain transmembrane #status predicted #label TM1\
33-49 #domain transmembrane #status predicted #label TM2\
52-68 #domain transmembrane #status predicted #label TM3\
95-111 #domain transmembrane #status predicted #label TM4\
138-154 #domain transmembrane #status predicted #label TM5\
206-222 #domain transmembrane #status predicted #label TM6\
237-253 #domain transmembrane #status predicted #label TM7\
289-305 #domain transmembrane #status predicted #label TM8\
310-326 #domain transmembrane #status predicted #label TM9\
378-394 #domain transmembrane #status predicted #label TM10\
422-438 #domain transmembrane #status predicted #label TM11\
464-480 #domain transmembrane #status predicted #label TM12
#length 487 #molecular-weight 52906 #checksum 1643
SUMMARY
Query Match 72.6%; Score 53; DB 1; Length 487;
Best Local Similarity 66.7%; Pred. No. 5.25e+00;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Db 165 IIPYIRNL 173
Qy 1 IIPYIVRKL 9
RESULT 6
ENTRY #type complete
TITLE hypothetical protein sir0818 - Synecocystis sp. (strain PCC
6803)
ORGANISM #formal_name Synecocystis sp.
#variety
DATE 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change
21-Aug-1998
ACCESSIONS S75569
REFERENCE S74322
#authors Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.;
Nakamura, Y.; Miyajima, N.; Hirose, M.; Sugliura, M.;
Sasamoto, S.; Kimura, T.; Hosouchi, T.; Matsuno, A.;
Muraki, A.; Nakazaki, N.; Naruo, K.; Okumura, S.; Shimpo,
S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.;
Yasuda, M.; Tabata, S.
DNA Res. (1996) 3:109-136
#journal Sequence analysis of the genome of the unicellular
cyanobacterium Synecocystis sp. PCC6803. II. Sequence
determination of the entire genome and assignment of
potential protein-coding regions.

```

```

#cross-references MUID:97061201
#accession S75569
#status preliminary
#molecule_type DNA
#residues 1-451 #label KAN
#cross-references EMBL:D90911; GB:AB001339; NID:gl653083; PID:dl018863;
PID:gl653214
#note the nucleotide sequence was submitted to the EMBL Data
Library, June 1996
SUMMARY #length 451 #molecular-weight 50417 #checksum 4508
Query Match 71.2%; Score 52; DB 2; Length 451;
Best Local Similarity 62.5%; Pred. No. 8.17e+00;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
Db 110 LPYPMVRR 117
Qy 1 IIPYIVRKL 8
RESULT 7
ENTRY #type complete
TITLE DNA-directed RNA polymerase (EC 2.7.7.6) chain A - Sulfolobus
acidocaldarius
ORGANISM #formal_name Sulfolobus acidocaldarius
DATE 09-Mar-1990 #sequence_revision 09-Mar-1990 #text_change
12-Sep-1997
ACCESSIONS B33926; S04717
REFERENCE A33926
#authors Puhler, G.; Leffers, H.; Gropp, F.; Palm, P.; Klenk, H.P.;
Lottspeich, F.; Garrett, R.A.; Zillig, W.
#journal Proc. Natl. Acad. Sci. U.S.A. (1989) 86:4569-4573
#title Archaeobacterial DNA-dependent RNA polymerases testify to the
evolution of the eukaryotic nuclear genome.
#cross-references MUID:89282812
#accession B33926
#status preliminary; nucleic acid sequence not shown; not
compared with conceptual translation
#molecule_type DNA
#residues 1-880 #label PUE
REFERENCE S04714
#authors Puhler, G.; Lottspeich, F.; Zillig, W.
#journal Nucleic Acids Res. (1989) 17:4517-4534
#title Organization and nucleotide sequence of the genes encoding
the large subunits A, B and C of the DNA-dependent RNA
polymerase of the archaeobacterium Sulfolobus
acidocaldarius
#cross-references MUID:89315197
#accession S04717
#molecule_type DNA
#residues 1-311, 'N', 313-560, 'N', 562-610, 'M', 612-640, 'M', 642-880
#label PUE
#cross-references EMBL:X14818; NID:gl66667; PID:gl66670
GENETICS
#gene rpoA
CLASSIFICATION #superfamily Halobacterium DNA-directed RNA polymerase chain
A
KEYWORDS nucleotidyltransferase; transcription
SUMMARY #length 880 #molecular-weight 99825 #checksum 9710
Query Match 71.2%; Score 52; DB 2; Length 880;
Best Local Similarity 66.7%; Pred. No. 8.17e+00;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
Db 339 IIPYIRNL 347
Qy 1 IIPYIVRKL 9
RESULT 8
ENTRY #type complete
TITLE hypothetical protein yopB - Bacillus subtilis
ORGANISM #formal_name Bacillus subtilis

```

DATE 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change  
24-Sep-1998  
ACCESSIONS H69915  
REFERENCE A69580  
#authors Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.;  
Alloni, G.; Azevedo, V.; Bertero, M.G.; Bessieres, P.;  
Bolotin, A.; Borchert, S.; Boriss, R.; Boursier, L.; Brans,  
A.; Braun, M.; Brignell, S.C.; Bron, S.; Brouillet, S.;  
Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;  
Choi, S.K.; Codani, J.J.; Connerton, I.F.; Cummings, N.J.;  
Daniel, R.A.; Denizot, F.; Devine, K.M.; Duesterhoeft, A.;  
Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.;  
Fabret, C.; Ferrari, E.; Foulger, D.; Fritz, C.; Fujita, M.;  
Fujita, Y.; Fuma, S.; Galluzzi, A.; Galleron, N.; Ghm,  
S.Y.; Glaser, P.; Goffeau, A.; Gollightly, E.J.; Grandi, G.;  
Guiseppi, G.; Guy, B.J.; Haga, K.; Halech, J.; Harwood,  
C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.;  
Hullo, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karamata, D.;  
Kashara, Y.; Kiehr-Blanchard, M.; Klein, C.; Kobayashi,  
Y.; Koetter, P.; Konigstein, G.; Krogh, S.; Kumano, M.;  
Kurita, K.; Lapidus, A.; Lardinois, S.; Lauber, J.;  
Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.;  
Mauel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno,  
M.; Mostl, D.; Nakai, S.; Noack, M.; Noone, D.; O'Reilly,  
M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro,  
V.; Pohl, T.M.; Portetelle, D.; Porwollik, S.; Prescott, A.M.;  
Presecan, E.; Pujic, P.; Purnelle, B.; Rapoport, G.;  
Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.;  
Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon, E.;  
Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.;  
Sekowska, A.; Seror, S.J.; Serron, P.; Shin, B.S.; Soldo,  
B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Takahashi, H.;  
Takemaru, K.; Takeuchi, M.; Tamakoshi, A.; Tanaka, T.;  
Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.;  
Vandenbol, M.; Vannier, F.; Vassarotti, A.; Viari, A.;  
Wambutt, R.; Wedler, E.; Wedler, H.; Weitzenecker, T.;  
Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto,  
K.; Yata, K.; Yoshida, K.; Yoshikawa, H.F.; Zumstein, E.;  
Yoshikawa, H.; Danchin, A.  
#journal Nature (1997) 390:249-256  
#title The complete genome sequence of the Gram-positive bacterium  
Bacillus subtilis.  
#cross-references MUID:98044033  
#accession H69915  
#status preliminary; nucleic acid sequence not shown;  
translation not shown  
#molecule\_type DNA  
#residues 1-75 #label KUN  
#cross-references GB:299115; GB:AL009126; NID:g2634478; PID:ell83542;  
PID:g2634515  
#experimental\_source strain 168  
GENETICS  
#gene yopB  
#summary #length 75 #molecular-weight 9099 #checksum 9896  
Query Match 69.9%; Score 51; DB 2; Length 75;  
Best Local Similarity 85.7%; Pred. No. 1.27e+01;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Db 28 YPIVRKL 34  
QY 3 YPIVRKL 9  
RESULT 9  
ENTRY C64422 #type complete  
TITLE hypothetical protein MJ0979 - Methanococcus jannaschii  
ORGANISM #formal\_name Methanococcus jannaschii  
DATE 13-Sep-1996 #sequence\_revision 13-Sep-1996 #text\_change  
10-Oct-1997  
ACCESSIONS C64422  
REFERENCE A64300  
#authors Buit, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann,

R.D.; Sutton, G.G.; Blake, J.A.; FitzGerald, L.M.; Clayton,  
R.A.; Gocayne, J.D.; Kerlavage, A.R.; Dougherty, B.A.;  
Tomb, J.F.; Adams, M.D.; Reich, C.I.; Overbeek, R.;  
Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.;  
Scott, J.L.; Geoghegan, N.S.M.; Weidman, J.F.; Fuhrmann,  
J.D.; Nguyen, D.; Utterback, T.R.; Kelley, J.M.; Peterson,  
J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts,  
K.M.; Hurst, M.A.; Kaine, B.P.; Borodovsky, M.; Klenk,  
H.P.; Fraser, C.M.; Smith, H.O.; Woese, C.R.; Venter, J.C.  
#journal Science (1996) 273:1058-1073  
#title Complete genome sequence of the methanogenic archaeon,  
Methanococcus jannaschii.  
#cross-references MUID:96537999  
#accession C64422  
#status preliminary; nucleic acid sequence not shown;  
translation not shown  
#molecule\_type DNA  
#residues 1-197 #label BUI  
#cross-references GB:U67541; GB:L77117; NID:gl591641; PID:gl499818;  
TIGR:MJ0979  
GENETICS  
#map\_position REV912311-911718  
#summary #length 197 #molecular-weight 21520 #checksum 5590  
Query Match 69.9%; Score 51; DB 2; Length 197;  
Best Local Similarity 55.6%; Pred. No. 1.27e+01;  
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;  
Db 189 IAPYIRKV 197  
QY 1 IPIPIVRKL 9  
RESULT 10  
ENTRY S62493 #type complete  
TITLE hypothetical protein SPAC23D3.02 - fission yeast  
ORGANISM #formal\_name Schizosaccharomyces pombe  
DATE 16-May-1996 #sequence\_revision 13-Mar-1997 #text\_change  
21-Aug-1998  
ACCESSIONS S62493  
REFERENCE S62492  
#authors Niblett, D.; Harris, D.  
#submission submitted to the EMBL Data Library, October 1995  
#accession S62493  
#status preliminary  
#molecule\_type DNA  
#residues 1-340 #label NIB  
#cross-references EMBL:264354; NID:gl039338; PID:gl039340  
GENETICS  
#map\_position 1R  
#introns 25/2  
CLASSIFICATION #superfamily phage T4 DNA polymerase accessory protein 44  
#summary #length 340 #molecular-weight 37876 #checksum 2036  
Query Match 69.9%; Score 51; DB 2; Length 340;  
Best Local Similarity 55.6%; Pred. No. 1.27e+01;  
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
Db 249 VPYNIRSL 257  
QY 1 IPIPIVRKL 9  
RESULT 11  
ENTRY F70350 #type complete  
TITLE recombination protein recN - Aquifex aeolicus  
ORGANISM #formal\_name Aquifex aeolicus  
DATE 08-May-1998 #sequence\_revision 08-May-1998 #text\_change  
08-May-1998  
ACCESSIONS F70350  
REFERENCE A70300  
#authors Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.;

```

Lenox, A.L.; Graham, D.E.; Overbeek, R.; Sneed, M.A.;
Keller, M.; AuJay, M.; Huber, R.; Feldman, R.A.; Short,
J.M.; Olson, G.J.; Swanson, R.V.
#journal Nature (1998) 392:353-358
#title The complete genome of the hyperthermophilic bacterium
Aquifex aeolicus.
#cross-references MUID:98196666
#accession F70350
##status preliminary; nucleic acid sequence not shown;
translation not shown
##molecule_type DNA
##residues 1-520 ##label AOF
##cross-references GB:AE000695; NID:g2983180; PID:g2983189; GB:AE000657
##experimental_source strain VF5
GENETICS
#gene recN
SUMMARY
#length 520 #molecular-weight 60439 #checksum 5555
Query Match 69.9%; Score 51; DB 2; Length 520;
Best Local Similarity 66.7%; Pred. No. 1.27e+01;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
Db 480 IPYIVREL 488
||| :|||
QY 1 IPYIVRKL 9
RESULT 12
ENTRY NUUTB #type complete
TITLE glucose-6-phosphate isomerase (EC 5.3.1.9) - Trypanosoma
brucei
ALTERNATE_NAMES phosphoglucose isomerase; phosphohexose isomerase
ORGANISM #formal_name Trypanosoma brucei
DATE 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change
05-Sep-1997
ACCESSIONS S06113
REFERENCE S06113
#authors Marchand, M.; Kooystra, U.; Wierenga, R.K.; Lambelir, A.M.;
van Beekun, J.; Opperdoes, F.R.; Michels, P.A.M.
#journal Eur. J. Biochem. (1989) 184:455-464
#title Glucosephosphate isomerase from Trypanosoma brucei. Cloning
and characterization of the gene and analysis of the
enzyme.
#cross-references MUID:90005496
#accession S06113
##molecule_type DNA
##residues 1-607 ##label MAR
#note part of this sequence was confirmed by protein
sequencing
CLASSIFICATION #superfamily glucose-6-phosphate isomerase
KEYWORDS gluconeogenesis; glycolysis; homodimer; intramolecular
oxidoreductase; isomerase
FEATURE
571 #active_site Lys #status predicted
SUMMARY #length 607 #molecular-weight 67517 #checksum 4076
Query Match 69.9%; Score 51; DB 1; Length 607;
Best Local Similarity 55.6%; Pred. No. 1.27e+01;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
Db 31 IPYEVTRRL 39
||| :|||
QY 1 IPYIVRKL 9
RESULT 13
ENTRY A56394 #type complete
TITLE pyocin S3 - Pseudomonas aeruginosa (strain P12)
ORGANISM #formal_name Pseudomonas aeruginosa
DATE 19-Oct-1995 #sequence_revision 19-Oct-1995 #text_change
09-Sep-1997
ACCESSIONS A56394

```

```

REFERENCE A56394
#authors Duport, C.; Bayse, C.; Michel-Briand, Y.
#journal J. Biol. Chem. (1995) 270:8920-8927
#title Molecular characterization of pyocin S3, a novel S-type
pyocin from Pseudomonas aeruginosa.
#cross-references MUID:95238389
#accession A56394
##status preliminary
##residues 1-766 ##label DUP
##cross-references GB:X77995; NID:g854362; PID:g854363
GENETICS
#gene pyoS3A
KEYWORDS bacteriocin
FEATURE
2-766 #product pyocin S3 #status experimental #label MAT
SUMMARY #length 766 #molecular-weight 81434 #checksum 2106
Query Match 69.9%; Score 51; DB 2; Length 766;
Best Local Similarity 66.7%; Pred. No. 1.27e+01;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
Db 679 IPYGEIRKL 687
||| :|||
QY 1 IPYPIVRKL 9
RESULT 14
ENTRY G70577 #type complete
TITLE probable dihydroorotate dehydrogenase - Mycobacterium
tuberculosis (strain H37RV)
ORGANISM #formal_name Mycobacterium tuberculosis
DATE 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change
17-Jul-1998
ACCESSIONS G70577
REFERENCE A70500
#authors Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher,
C.; Harris, D.; Gordon, S.V.; Eiglmeier, K.; Gas, S.; Barry
III, C.E.; Tekala, F.; Badcock, K.; Basham, D.; Brown, D.;
Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.;
Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;
Hornsby, T.; Jagels, K.; Krogh, A.; McLean, J.; Moule, S.;
Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.;
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.;
Skellton, S.; Squares, S.; Squires, R.; Sulston, J.E.;
Taylor, K.; Whitehead, S.; Barrett, B.G.
#journal Nature (1998) 393:537-544
#title Deciphering the biology of Mycobacterium tuberculosis from
the complete genome sequence.
#cross-references MUID:98295987
#accession G70577
##status preliminary; nucleic acid sequence not shown;
translation not shown
##molecule_type DNA
##residues 1-357 ##label COL
##cross-references GB:Z95388; GB:AL123456; NID:g3261759; PID:e316034;
PID:g2104339
##experimental_source strain H37RV
GENETICS
#gene pyrD
SUMMARY #length 357 #molecular-weight 37998 #checksum 1275
Query Match 68.5%; Score 50; DB 2; Length 357;
Best Local Similarity 71.4%; Pred. No. 1.95e+01;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
Db 2 YPLVRRRL 8
||| :|||
QY 3 YPIVRKL 9
RESULT 15
ENTRY H64769 #type complete

```

preprotein translocase chain secD - Escherichia coli  
 TITLE protein-export membrane protein secD; secretion protein secD  
 ALTERNATE\_NAMES #formal\_name Escherichia coli  
 ORGANISM 12-Sep-1997 #sequence\_revision 17-Sep-1997 #text\_change  
 DATE 12-Feb-1999  
 ACCESSIONS H64769; JQ0696; S12301  
 REFERENCE A64720  
 #authors Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.;  
 Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.;  
 Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.;  
 Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao,  
 Y.  
 #journal Science (1997) 277:1453-1462  
 #title The complete genome sequence of Escherichia coli K-12.  
 #cross-references MUID:97426617  
 #accession H64769  
 #status nucleic acid sequence not shown; translation not shown  
 #molecule\_type DNA  
 #residues 1-615 #label BLAT  
 #cross-references GB:AF000147; GB:U00096; NID:g1786603; PID:g1786609;  
 UWGP:b0408  
 #experimental\_source strain K-12, substrain MG1655  
 REFERENCE JQ0693  
 #authors Gardel, C.; Johnson, K.; Jacq, A.; Beckwith, J.  
 #journal EMBO J. (1990) 9:3209-3216  
 #title The secD locus of E.coli codes for two membrane proteins  
 #cross-references MUID:91006014  
 #accession JQ0696  
 #molecule\_type DNA  
 #residues 1-77, 'S', '79-154', 'A', '156-615' #label GAR  
 #cross-references GB:X56175; NID:g42929; PID:g581230  
 REFERENCE S12298  
 #authors Gardel, C.; Johnson, K.; Jacq, A.; Beckwith, J.  
 #journal EMBO J. (1990) 9:4205-4206  
 #contents erratum  
 #accession S12301  
 #molecule\_type DNA  
 #residues 1-77, 'S', '79-154', 'A', '156-615' #label GA2  
 #cross-references EMBL:X56175; NID:g42929; PID:g581230  
 REFERENCE A36969  
 #authors Pogliano, K.J.; Beckwith, J.  
 #journal J. Bacteriol. (1994) 176:804-814  
 #title Genetic and molecular characterization of the Escherichia  
 coli secD operon and its products.  
 #contents annotation; membrane topology  
 COMMENT Preprotein translocase contains a membrane-embedded trimeric  
 complex of SecY, SecE and SecZ and the peripheral SecA protein.  
 The proteins SecD, SecF and YajC also form an integral membrane  
 heterotrimeric complex. These two trimeric complexes are  
 associated to form SecYEGFYajC, the hexameric integral membrane  
 domain of the pre- protein translocase 'holoenzyme'.  
 GENETICS  
 #gene secD  
 #start\_codon GTG  
 COMPLEX heterohexameric; chains secY, secE, secG, secD, secF, and yajC  
 CLASSIFICATION #superfamily protein export membrane protein secD  
 KEYWORDS inner membrane; protein export; transmembrane protein  
 FEATURE  
 10-30 #domain transmembrane #status predicted #label TM1\  
 31-455 #domain periplasmic #status predicted #label PF1\  
 456-472 #domain transmembrane #status predicted #label TM2\  
 477-497 #domain transmembrane #status predicted #label TM3\  
 498-501 #domain periplasmic #status predicted #label PP2\  
 502-518 #domain transmembrane #status predicted #label TM4\  
 564-580 #domain transmembrane #status predicted #label TM5\  
 581-585 #domain periplasmic #status predicted #label PF3\  
 586-605 #domain transmembrane #status predicted #label TM6  
 #length 615 #molecular-weight 66631 #checksum 9609  
 SUMMARY  
 Query Match 68.5%; Score 50; DB 2; Length 615;  
 Best Local Similarity 75.0%; Pred. No. 1.95e+01;  
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 163 IPYTTVRK 170  
 ||| |||  
 QY 1 IPYPIVRK 8  
 Search completed: Fri Apr 14 23:43:27 2000  
 Job time : 9 secs.



\*\*\*\*\*  
[M][A][P][S][E][R][C][H] (TM)  
\*\*\*\*\*

Release 3.1A John F. Collins, BioComputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Apr 14 23:43:45 2000; MasPar time 5.65 Seconds  
Tabular output not generated. 47.576 Million cell updates/sec

Title: >US-08-452-843-8  
Description: (1-9) from US08452843.pep  
Perfect Score: 73  
Sequence: 1 IPYPIVRKL 9

Scoring table: PAM 150  
Gap 15

Searched: 82229 seqs, 29864866 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: swiss-prot38  
1:swissprot

Statistics: Mean 25.156; Variance 27.317; scale 0.921

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	55	75.3	1679	1	YIO9_YEAST HYPOTHETICAL 195.1 KD	4.63e+01
2	53	72.6	209	1	RAS_GEOCY RAS-LIKE PROTEIN.	1.32e+00
3	53	72.6	487	1	TTDT_ECOLI PUTATIVE TARTRATE CARR	1.32e+00
4	53	72.6	1033	1	YDK9_SCHPO HYPOTHETICAL 116.5 KD	1.32e+00
5	52	71.2	880	1	RPAL_SULAC DNA-DIRECTED RNA POLYM	2.21e+00
6	51	69.9	133	1	SY21_MOUSE SMALL INDUCIBLE CYTOKI	3.66e+00
7	51	69.9	197	1	Y979_METJA HYPOTHETICAL PROTEIN M	3.66e+00
8	51	69.9	340	1	RFC2_SCHPO PROBABLE ACTIVATOR 1 4	3.66e+00
9	51	69.9	607	1	G6PL_TRYBB GLUCOSE-6-PHOSPHATE IS	3.66e+00
10	51	69.9	824	1	DPOL_METVO DNA POLYMERASE (EC 2.7	3.66e+00
11	50	68.5	356	1	OXDA_TRIVR D-AMINO ACID OXIDASE (	6.02e+00
12	50	68.5	357	1	PYRD_MYCTU DIHYDROOATATE DEHYDRO	6.02e+00
13	50	68.5	615	1	SECD_SALCH PROTEIN-EXPORT MEMBRAN	6.02e+00
14	50	68.5	615	1	SECD_ECOLI PROTEIN-EXPORT MEMBRAN	6.02e+00
15	49	67.1	183	1	APT_METJA ADENINE PHOSPHORIBOSYL	9.82e+00
16	49	67.1	342	1	ARGI_ARATH ARGINASE (EC 3.5.3.1).	9.82e+00
17	49	67.1	350	1	ARGI_SOYBN ARGINASE (EC 3.5.3.1).	9.82e+00
18	49	67.1	1061	1	RNE_ECOLI RIBONUCLEASE E (EC 3.1	9.82e+00
19	49	67.1	1676	1	CO5_HUMAN COMPLEMENT C5 PRECURSO	9.82e+00
20	49	67.1	1680	1	CO5_MOUSE COMPLEMENT C5 PRECURSO	9.82e+00
21	48	65.8	100	1	YOGA_CAEEL HYPOTHETICAL 11.2 KD P	1.59e+01
22	48	65.8	213	1	AMIS_MYCSM PUTATIVE AMIDATE SUBST	1.59e+01
23	48	65.8	305	1	LIGD_PSEPA C ALPHA-DEHYDROGENASE	1.59e+01

24	48	65.8	311	1	HTRB_HAEIN LIPID A BIOSYNTHESIS L	1.59e+01
25	48	65.8	326	1	GLN2_RHILP GLUTAMINE SYNTHETASE I	1.59e+01
26	48	65.8	356	1	YDGC_SCHPO HYPOTHETICAL 41.3 KD P	1.59e+01
27	48	65.8	372	1	DP3B_CAUCR DNA POLYMERASE III, BE	1.59e+01
28	48	65.8	377	1	TRA7_BACST PUTATIVE TRANSPOSASE F	1.59e+01
29	48	65.8	407	1	CPXD_AGR75 CYTOCHROME P450-P1NF2,	1.59e+01
30	48	65.8	540	1	TH16_YEAST THIAMINE BIOSYNTHETIC	1.59e+01
31	48	65.8	617	1	VGF_RAT VGF PROTEIN PRECURSOR	1.59e+01
32	48	65.8	762	1	TRPG_NEUCR ANTHRANILATE SYNTHASE	1.59e+01
33	47	64.4	196	1	APT_METJA ADENINE PHOSPHORIBOSYL	2.55e+01
34	47	64.4	238	1	Y106_METJA HYPOTHETICAL PROTEIN M	2.55e+01
35	47	64.4	268	1	YWDF_BACSU HYPOTHETICAL 30.6 KD P	2.55e+01
36	47	64.4	323	1	YMC2_SCHPO HYPOTHETICAL COX1 INTR	2.55e+01
37	47	64.4	329	1	I329_ASFB7 LATE PROTEIN I329L PRE	2.55e+01
38	47	64.4	329	1	GLN2_BRAJA GLUTAMINE SYNTHETASE I	2.55e+01
39	47	64.4	381	1	TH11_METJA PROBABLE THIAMINE BIOS	2.55e+01
40	47	64.4	472	1	YAE3_SCHPO HYPOTHETICAL 54.3 KD P	2.55e+01
41	47	64.4	562	1	GR78_NEUCR 78 KD GLUCOSE-REGULATE	2.55e+01
42	47	64.4	582	1	GR78_YEAST 78 KD GLUCOSE-REGULATE	2.55e+01
43	47	64.4	718	1	PLSB_CAEEL PROBABLE GLYCEROL-3-PH	2.55e+01
44	47	64.4	1056	1	YNN2_YEAST HYPOTHETICAL 119.3 KD	2.55e+01
45	47	64.4	1868	1	YHD0_YEAST HYPOTHETICAL 210.4 KD	2.55e+01

ALIGNMENTS

RESULT 1  
ID YIO9\_YEAST STANDARD; PRT; 1679 AA.  
AC P40457.  
DT 01-FEB-1995 (Rel. 31, Created)  
DT 01-FEB-1995 (Rel. 31, Last sequence update)  
DE 01-FEB-1995 (Rel. 31, Last annotation update)  
DE HYPOTHETICAL 195.1 KD PROTEIN IN DNA43-UBI1 INTERGENIC REGION.  
GN YII149C.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;  
OC Saccharomycetaceae; Saccharomycetes.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=S288C / AB972;  
RA BARRELL B.G., BADCOCK K., BANKIER A.T., BOWMAN S., BROWN D.,  
RA CHURCHER C.M., CONNOR R., COPSEY T., DEAR S., DEVLIN K., FRASER A.,  
RA GENTLES S., HAMLYN N., HORSNELL T.S., HUNT S., JAGELS K., JONES M.,  
RA LOUIS E., LYE G., MOULE S., MOULE C., ODELL C., PEARSON D.,  
RA RAJANDREAM M.A., RILES L., ROWLEY N., SKELTON J., SMITH V.,  
RA WALSH S.V., WHITEHEAD S.;  
RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC -----  
CC EMBL; 238059; CAA86129.1; -.  
DR PIR; S48385; S48385.  
KW Hypothetical protein.  
SQ SEQUENCE 1679 AA; 195141 MW; 5897CD94 CRC32;

Query Match 75.3%; Score 55; DB 1; Length 1679;  
Best Local Similarity 55.6%; Pred. No. 4.63e+01;  
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;  
Db 19 VTIPIVRKL 27  
QY : ||:|||  
1 IPYPIVRKL 9  
RESULT 2  
ID RAS\_GEOCY STANDARD; PRT; 209 AA.  
AC P24498;



```
Db 165 IYPIRNL 173
   |||||:|
Qy 1 IYPIVRKL 9

RESULT 4
ID YDK9_SCHPO STANDARD; PRT; 1033 AA.
AC P87115;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE HYPOTHETICAL 116.5 KD PROTEIN C2068.09C IN CHROMOSOME I.
GN SPAC2068.09C.
OS Schizosaccharomyces pombe (Fission Yeast).
OC Eukaryota; Fungi; Ascomycota; Archiascomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-972;
RA BADCOCK K., CHURCHER C.M., WOOD V., BARRELL B.G., RAVANDREEM M.A.;
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: TO YEAST YNL132W AND AN A.AMBISEXUALIS HYPOTHETICAL
CC PROTEIN (AC P54008).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X14818; CAA32925.1; -
CC PIR; S04717; S04717.
CC PFAM; PF00623; RNA_pol_A; 1.
CC Transferase; DNA-directed RNA polymerase; Transcription; Zinc.
FT ZN_FING 58 101 POTENTIAL.
SQ SEQUENCE 880 AA; 99790 MW; 665B33F9 CRC32;

Query Match 71.2%; Score 52; DB 1; Length 880;
Best Local Similarity 66.7%; Pred. No. 2.21e+00;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 339 IYPIARML 347
   |||||:|
Qy 1 IYPIVRKL 9

RESULT 6
ID SV21_MOUSE STANDARD; PRT; 133 AA.
AC O09006; O09002;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE SMALL INDUCIBLE CYTOKINE A21 PRECURSOR (BETA CHEMOKINE EXODUS-2)
DE (6CKINE) (THYMUS-DERIVED CHEMOTACTIC AGENT 4) (TCA4).
GN SCYA21.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-FETAL;
RX MEDLINE; 97444139.
RA HROMAS R.A., KIM C.H., KLEMSZ M., KRATHWOHL M., FIFE K., COOPER S.,
RA SCHNIZLEIN-BICK C., BROXMEYER H.E.;
RT "Isolation and characterization of Exodus-2, a novel C-C chemokine
RT with a unique 37-amino acid carboxyl-terminal extension.";
RL J. Immunol. 159:2554-2558(1997).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 97400322.
RA HEDRICK J.A., ZLOTNIK A.;
RT "Identification and characterization of a novel beta chemokine
RT containing six conserved cysteines.";
RL J. Immunol. 159:1589-1593(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE-THYMUS;
RX MEDLINE; 98208291.
RA TANABE S., LU Z., LUO Y., QUACKENBUSH E.J., BERMAN M.A.,
RA COLLINS-RACIE L.A., MI S., REILLY C., LO D., JACOBS K.A., DORF M.E.;
RT "Identification of a new mouse beta-chemokine, thymus-derived
RT chemotactic agent 4, with activity on T lymphocytes and mesangial
RT cells.";
RL J. Immunol. 159:5671-5679(1997).
CC -1- FUNCTION: INHIBITS HEMOPOIESIS AND STIMULATE CHEMOTAXIS.
CC CHEMOTACTIC IN VITRO FOR THYMOCYTES AND ACTIVATED T CELLS, BUT NOT
CC FOR B CELLS, MACROPHAGES, OR NEUTROPHILS.
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- TISSUE SPECIFICITY: BROADLY EXPRESSED; FOUND IN SPLEEN AND LUNG.
CC LYMPHOID ORGANS, PARTICULARLY LYMPH NODE, SPLEEN, AND APPENDIX.
CC -1- SIMILARITY: BELONGS TO THE INTERCRINE BETA FAMILY (SMALL CYTOKINE
```

```

DR ENBL; U67541; AAB98994.1; -.
DR TIGR; M00979; -.
KW Hypothetical protein; Transmembrane.
FT TRANSMEM 11 31 POTENTIAL.
FT TRANSMEM 85 105 POTENTIAL.
FT TRANSMEM 109 129 POTENTIAL.
FT TRANSMEM 174 194 POTENTIAL.
SQ SEQUENCE 197 AA; 21520 MW; 926EC4E9 CRC32;

Query Match          69.9%  Score 51;  DB 1;  Length 197;
Best Local Similarity 55.6%  Pred. No. 3.66e+00;
Matches 5;  Conservative 4;  Mismatches 0;  Indels 0;  Gaps 0;

Db 189 IAYPIRKV 197
   |||:|
QY 1 IYPIVRKL 9

RESULT 8
ID REC2_SCHPO STANDARD; PRT; 340 AA.
AC Q09843;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE PROBABLE ACTIVATOR 1 41 KD SUBUNIT (REPLICATION FACTOR C 41 KD
SUBUNIT).
DE SPAC23D3.02.
GN Schizosaccharomyces pombe (Fission yeast).
OS Eukaryota; Fungi; Ascomycota; Archiascomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
CC Schizosaccharomycetes.
CC [1]
RN SEQUENCE FROM N.A.
RP STRAIN-972;
RC NIELETT D., HARRIS D., BARRELL B.G., RAJANDREAM M.A., WALSH S.V.;
RL submitted (OCT-1995) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: THE ELONGATION OF PRIMED DNA TEMPLATES BY DNA POLYMERASE
DELTA AND EPSILON REQUIRES THE ACTION OF THE ACCESSORY PROTEINS
PROLIFERATING CELL NUCLEAR ANTIGEN (PCNA) AND ACTIVATOR 1. THE
41 KD SUBUNIT BINDS ATP AND TO SINGLE-STRANDED DNA
(BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE ACTIVATOR 1 36 TO 40 KD SUBUNITS
FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announcement/
or send an email to license@isb-sib.ch).
-----
CC ENBL; 264354; CAA91237.1;
DR Hypothetical protein; DNA replication; ATP-binding; Nuclear protein;
KW DNA-binding.
FT NP_BIND 59 66 ATP (POTENTIAL).
SQ SEQUENCE 340 AA; 37876 MW; FB518443 CRC32;

Query Match          69.9%  Score 51;  DB 1;  Length 340;
Best Local Similarity 55.6%  Pred. No. 3.66e+00;
Matches 5;  Conservative 2;  Mismatches 0;  Indels 0;  Gaps 0;

Db 249 VPMNIIRSL 257
   |||:|
QY 1 IYPIVRKL 9

RESULT 9
ID G6FI_TRYEBB STANDARD; PRT; 607 AA.
AC P13377;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)

```

```
DT 01-DEC-1992 (Rel. 24, Last annotation update)
DE GLUCOSE-6-PHOSPHATE ISOMERASE, GLYCOSOMAL (GPI) (EC 5.3.1.9)
DE (PHOSPHOGLUCOSE ISOMERASE) (PGI) (PHOSPHOHEXOSE ISOMERASE) (PHI).
GN PGI.
OS Trypanosoma brucei brucei.
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
RN [1]
RN SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=427;
RX MEDLINE: 90005496.
RA MARCHAND M., KOOSYRA U., WIERENGA R.K., LAMBEIR A.M., VAN BEEDUMEN J.,
RA OPPERDOES F.R., MICHELS P.A.M.;
RT "Glucosephosphate isomerase from Trypanosoma brucei. Cloning and
RT characterization of the gene and analysis of the enzyme.";
RL Eur. J. Biochem. 184:455-464(1989).
CC -1- CATALYTIC ACTIVITY: GLUCOSE 6-PHOSPHATE = FRUCTOSE 6-PHOSPHATE.
CC -1- PATHWAY: INVOLVED IN GLYCOLYSIS AND IN GLUCONEOGENESIS.
CC -1- SUBUNIT: HOMODIMER.
CC -1- SUBCELLULAR LOCATION: GLYCOSOMAL.
CC -1- SIMILARITY: BELONGS TO THE GPI FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X15540; CAA33547.1; -.
DR PIR: S06113; NUOTB.
DR PROSITE: PS00765; P_GLUCOSE_ISOMERASE_1; 1.
DR PROSITE: PS00174; P_GLUCOSE_ISOMERASE_2; 1.
DR PROSITE: PS00342; MICROBODIES_CTER; 1.
DR PFAM: PF00342; PGI; 1.
KW Glucosoneogenesis; Glycolysis; Isomerase; Glycosome.
FT SITE 605 607 MICROBODY TARGETING SIGNAL (POTENTIAL).
SQ SEQUENCE 607 AA; 67518 MW; EF35CC43 CRC32;

Query Match 69.9%; Score 51; DB 1; Length 607;
Best Local Similarity 55.6%; Pred. No. 3.66e+00;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 31 IPYEVTRKL 39
   ||| : | : |
QY 1 IPYIVRKL 9

RESULT 10
ID DPOL_METVO STANDARD; PRT; 824 AA.
AC P52025;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE DNA POLYMERASE (EC 2.7.7.7).
GN POL.
OS Methanococcus voltae.
OC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
OC Methanococcus.
RN [1]
RN SEQUENCE FROM N.A.
RX MEDLINE: 95014087.
RA KONISKY J., PAULE S.M., CARINATO M.E., KANSY J.W.;
RT "The DNA polymerase gene from the methanogenic archaeon Methanococcus
RT voltae.";
RL J. Bacteriol. 176:6402-6403(1994).
CC -1- CATALYTIC ACTIVITY: N DEOXYNUCLEOSIDE TRIPHOSPHATE =
CC N PYROPHOSPHATE + DNA(N).
CC -1- SIMILARITY: BELONGS TO DNA POLYMERASE TYPE-B FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L33366; AAA72443.1; -.
DR PROSITE: PS00116; DNA_POLYMERASE_B; 1.
DR PFAM: PF00136; DNA_pol_B; 3.
KW Transferase; DNA-directed DNA polymerase; DNA replication;
KW DNA-binding.
SQ SEQUENCE 824 AA; 96754 MW; 94579170 CRC32;

Query Match 69.9%; Score 51; DB 1; Length 824;
Best Local Similarity 85.7%; Pred. No. 3.66e+00;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 308 YPIARKL 314
   ||| |||
QY 3 YPIVRKL 9

RESULT 11
ID OXDA_TRIVR STANDARD; PRT; 356 AA.
AC Q39042;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE D-AMINO ACID OXIDASE (EC 1.4.3.3) (DAMOX) (DAO) (DAAO).
GN DAO1.
OS Trigonopsis variabilis.
OC Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
OC Candidaceae; Trigonopsis.
RN [1]
RN SEQUENCE FROM N.A.
RX STRAIN=CBS 4095;
RX MEDLINE: 98095789.
RA GONZALEZ F.J., MONTES J., MARTIN F., LOPEZ M.C., FERMINAN E.,
RA CATALAN J., GALAN M.A., DOMINGUEZ A.;
RT "Molecular cloning of TvDAO1, a gene encoding a D-amino acid oxidase
RT from Trigonopsis variabilis and its expression in Saccharomyces
RT cerevisiae and Kluyveromyces lactis.";
RL Yeast 13:1399-1408(1997).
CC -1- CATALYTIC ACTIVITY: A D-AMINO ACID + H(2)O + O(2) = A 2-OXO-ACID +
CC NH(3) + H(2)O(2).
CC -1- COFACTOR: FAD.
CC -1- SIMILARITY: BELONGS TO THE DAMOX/DASOX FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z50019; CAA90322.1; -.
DR PROSITE: PS00677; DAO; 1.
DR PFAM: PF01266; DAO; 1.
KW Oxidoreductase; Flavoprotein; FAD.
FT NP_BIND 4 18 FAD (ADP PART) (POTENTIAL).
FT ACT_SITE 243 243 BY SIMILARITY.
FT ACT_SITE 324 324 BY SIMILARITY.
SQ SEQUENCE 356 AA; 39301 MW; BA069642 CRC32;

Query Match 68.5%; Score 50; DB 1; Length 356;
Best Local Similarity 55.6%; Pred. No. 6.02e+00;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 66 VSPILREL 74
   : ||| : |
QY 1 IPYIVRKL 9
```

```

RESULT 12
ID PYRD MYCTU STANDARD; PRT; 357 AA.
AC O06236;
DT 15-DEC-1999 (Rel. 39, Created)
DT 15-DEC-1999 (Rel. 39, Last sequence update)
DE DIHYDROOROTATE DEHYDROGENASE (EC 1.3.3.1) (DIHYDROOROTATE OXIDASE)
DE (DHODEHASE).
GN PYRD OR RV2139 OR MTCY270.29C.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-H37RV;
RX MEDLINE; 98295987
RA COLE S.T., BROSCHE R., PARKHILL J., GARNIER T., CHURCHER C., HARRIS D.,
RA GORDON S.V., EIGLMEIER K., GAS S., BARRY C.E. III, TERAKA F.,
RA BADCOCK K., BASHAM D., BROWN D., CHILLINGWORTH T., CONNOR R.,
RA DAVIES R., DEVLIN K., FELTWEILL T., GENTILES S., HAMLIN N., HOLROYD S.,
RA HORNSBY T., JAGELS K., KROGH A., MCLEAN J., MOULE S., MURPHY L.,
RA OLIVER S., OSBORNE J., QUAIL M.A., RAJANDREAM M.A., ROGERS J.,
RA RUTHER S., SEGER K., SHELTON S., SQUARES S., SQUARES R., SULSTON J.E.,
RA TAYLOR K., WHITEHEAD S., BARRELL B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544(1998).
CC -1- CATALYTIC ACTIVITY: L-DIHYDROOROTATE + O(2) = OROTATE + H(2)O(2).
CC -1- COFACTOR: FAD (BY SIMILARITY).
CC -1- PATHWAY: FOURTH STEP IN PYRIMIDINE BIOSYNTHESIS.
CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: INNER SIDE OF THE MEMBRANE (BY SIMILARITY).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; Z95388; CAB08654.1;
DR PROSITE; PS00911; DHODEHASE_1; 1.
DR PROSITE; PS00912; DHODEHASE_2; 1.
DR PFAM; PF01180; DHODEHASE; 1.
KW Pyrimidine biosynthesis; Oxidoreductase; Flavoprotein; FAD.
FT NP_BIND 286 294
FT SEQUENCE 357 AA; 37998 MW; 56358C06 CRC32;
SQ
Query Match 68.5%; Score 50; DB 1; Length 357;
Best Local Similarity 71.4%; Pred. No. 6.02e+00;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 2 YPLVRL 8
QY 3 YPIVRL 9

RESULT 13
ID SECD_SALCH STANDARD; PRT; 615 AA.
AC Q92FF8;
DT 15-DEC-1999 (Rel. 39, Created)
DT 15-DEC-1999 (Rel. 39, Last sequence update)
DE DIHYDROOROTATE DEHYDROGENASE (EC 1.3.3.1) (DIHYDROOROTATE OXIDASE)
DE (DHODEHASE).
GN PYRD OR RV2139 OR MTCY270.29C.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-H37RV;
RX MEDLINE; 98295987
RA COLE S.T., BROSCHE R., PARKHILL J., GARNIER T., CHURCHER C., HARRIS D.,
RA GORDON S.V., EIGLMEIER K., GAS S., BARRY C.E. III, TERAKA F.,
RA BADCOCK K., BASHAM D., BROWN D., CHILLINGWORTH T., CONNOR R.,
RA DAVIES R., DEVLIN K., FELTWEILL T., GENTILES S., HAMLIN N., HOLROYD S.,
RA HORNSBY T., JAGELS K., KROGH A., MCLEAN J., MOULE S., MURPHY L.,
RA OLIVER S., OSBORNE J., QUAIL M.A., RAJANDREAM M.A., ROGERS J.,
RA RUTHER S., SEGER K., SHELTON S., SQUARES S., SQUARES R., SULSTON J.E.,
RA TAYLOR K., WHITEHEAD S., BARRELL B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544(1998).
CC -1- CATALYTIC ACTIVITY: L-DIHYDROOROTATE + O(2) = OROTATE + H(2)O(2).
CC -1- COFACTOR: FAD (BY SIMILARITY).
CC -1- PATHWAY: FOURTH STEP IN PYRIMIDINE BIOSYNTHESIS.
CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: INNER SIDE OF THE MEMBRANE (BY SIMILARITY).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; Z95388; CAB08654.1;
DR PROSITE; PS00911; DHODEHASE_1; 1.
DR PROSITE; PS00912; DHODEHASE_2; 1.
DR PFAM; PF01180; DHODEHASE; 1.
KW Pyrimidine biosynthesis; Oxidoreductase; Flavoprotein; FAD.
FT NP_BIND 286 294
FT SEQUENCE 357 AA; 37998 MW; 56358C06 CRC32;
SQ
Query Match 68.5%; Score 50; DB 1; Length 357;
Best Local Similarity 71.4%; Pred. No. 6.02e+00;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 2 YPLVRL 8
QY 3 YPIVRL 9

RESULT 14
ID SECD_ECOLI STANDARD; PRT; 615 AA.
AC P19673; P77531; P72348;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-DEC-1999 (Rel. 39, Last annotation update)
DE PROTEIN-EXPORT MEMBRANE PROTEIN SECD.
GN SECD.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RX MEDLINE; 9106014.
RA GARDEL C., JOHNSON K., JACQ A., BECKWITH J.;
RT "The secD locus of E.coli codes for two membrane proteins required
RT for protein export."
RL EMBO J. 9:3209-3216(1990).
RN [2]
RP ERRATUM.
RX MEDLINE; 91065354.
RA GARDEL C., JOHNSON K., JACQ A., BECKWITH J.;
RL EMBO J. 9:4205-4206(1990).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE; 97426617.
RA BLATTNER F.R., PLUNKETT G. III, BLOCH C.A., PERNA N.T., BURLAND V.,
RA RILEY M., COLLADO-VIDES J., GLASNER F.D., RODE C.K., MAYHEW G.F.,
RA GREGOR J., DAVIS N.W., KIRKPATRICK H.A., GOEDEN M.A., ROSE D.J.,
RA MAU B., SHAO Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
RN [4]
RP SEQUENCE FROM N.A.
RA ROBERTS D., ALLEN E., ARAUTO R., APARICIO A., CHUNG E., DAVIS K.,
RA DUNCAN M., FEDERSPIEL N., HYMAN R., KALMAN S., KOMP C., KURDI O.,
RA LEW H., LIN D., NAMATH A., OEFNER P., SCHRAMM S., DAVIS R.W.;
RT "Identification of secD gene from Salmonella choleraesuis."

```

```

Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
-1- FUNCTION: INVOLVED IN PROTEIN EXPORT.
-1- SUBUNIT: PART OF THE PROKARYOTIC PROTEIN TRANSLOCATION APPARATUS
WHICH COMPRISE SECA, SECB, SECD, SECE, SECF, SECG AND SECH
(BY SIMILARITY).
-1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE.
-1- SIMILARITY: BELONGS TO THE SECD/SECF FAMILY. SECD FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; AF100611; AAC83834.1;
KW Protein transport; Translocation; Transmembrane; Inner membrane.
FT TRANSMEM 10 30
FT TRANSMEM 452 472
FT TRANSMEM 504 524
FT TRANSMEM 564 584
FT SEQUENCE 615 AA; 66632 MW; 5BBEAEEL CRC32;
SQ
Query Match 68.5%; Score 50; DB 1; Length 615;
Best Local Similarity 75.0%; Pred. No. 6.02e+00;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 163 IPYTVRK 170
QY 1 IPYPIVRK 8

RESULT 14
ID SECD_ECOLI STANDARD; PRT; 615 AA.
AC P19673; P77531; P72348;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-DEC-1999 (Rel. 39, Last annotation update)
DE PROTEIN-EXPORT MEMBRANE PROTEIN SECD.
GN SECD.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RX MEDLINE; 9106014.
RA GARDEL C., JOHNSON K., JACQ A., BECKWITH J.;
RT "The secD locus of E.coli codes for two membrane proteins required
RT for protein export."
RL EMBO J. 9:3209-3216(1990).
RN [2]
RP ERRATUM.
RX MEDLINE; 91065354.
RA GARDEL C., JOHNSON K., JACQ A., BECKWITH J.;
RL EMBO J. 9:4205-4206(1990).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE; 97426617.
RA BLATTNER F.R., PLUNKETT G. III, BLOCH C.A., PERNA N.T., BURLAND V.,
RA RILEY M., COLLADO-VIDES J., GLASNER F.D., RODE C.K., MAYHEW G.F.,
RA GREGOR J., DAVIS N.W., KIRKPATRICK H.A., GOEDEN M.A., ROSE D.J.,
RA MAU B., SHAO Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
RN [4]
RP SEQUENCE FROM N.A.
RA ROBERTS D., ALLEN E., ARAUTO R., APARICIO A., CHUNG E., DAVIS K.,
RA DUNCAN M., FEDERSPIEL N., HYMAN R., KALMAN S., KOMP C., KURDI O.,
RA LEW H., LIN D., NAMATH A., OEFNER P., SCHRAMM S., DAVIS R.W.;
RT "Identification of secD gene from Salmonella choleraesuis."

```

```

[5]
RN  SEQUENCE OF 1-76 FROM N.A.
RP  MEDLINE; 94131960.
RA  POGLIANO K.J., BECKWITH J.;
RT  "Genetic and molecular characterization of the Escherichia coli secD
RL  operon and its products.";
RL  J. Bacteriol. 176:804-814(1994).
CC  -1- FUNCTION: INVOLVED IN PROTEIN EXPORT.
CC  -1- SUBUNIT: PART OF THE PROKARYOTIC PROTEIN TRANSLOCATION APPARATUS
CC  WHICH COMPRISE SECA, SECB, SECD, SECE, SECF, SECG AND SECH.
CC  -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE.
CC  -1- SIMILARITY: BELONGS TO THE SECY/SECF FAMILY. SECY FAMILY.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL; X56175; CAA39634.1; -.
DR  EMBL; AE000147; AAC73511.1; -.
DR  EMBL; U82664; AAB40164.1; -.
DR  EMBL; S68715; AAC60469.1; -.
DR  PIR; JQ0696; JQ0696.
DR  PIR; S12301; S12301.
DR  ECGENE; EG10938; SECY.
KW  Protein transport; Translocation; Transmembrane; Inner membrane.
FT  TRANSMEM 10 30 POTENTIAL.
FT  TRANSMEM 452 472 POTENTIAL.
FT  TRANSMEM 504 524 POTENTIAL.
FT  TRANSMEM 564 584 POTENTIAL.
FT  CONFLICT 78 78 F -> S (IN REF. 1).
FT  CONFLICT 155 155 R -> A (IN REF. 1).
SQ  SEQUENCE 615 AA; 66632 MW; 9943E19B CRC32;

Query Match 68.5%; Score 50; DB 1; Length 615;
Best Local Similarity 75.0%; Pred. No. 6.02e+00;
Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 163 IPYTVRK 170
   ||| |||
Qy 1 IPYPIVRK 8

RESULT 15
ID  APT_METJA STANDARD; PRT; 183 AA.
AC  Q59049;
DT  01-NOV-1997 (Rel. 35, Created)
DT  01-NOV-1997 (Rel. 35, Last sequence update)
DT  15-DEC-1999 (Rel. 39, Last annotation update)
DE  ADENINE PHOSPHORIBOSYLTRANSFERASE (EC 2.4.2.7) (APRT).
GN  APT OR MJ1655.
OS  Methanococcus jannaschii.
OC  Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
OC  Methanococcus.
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX  MEDLINE; 96337999.
RA  BULT C.J., WHITE O., OLSEN G.J., ZHOU L., FLEISCHMANN R.D.,
RA  SUTTON G.G., BLAKE J.A., FITZGERALD L.M., CLAYTON R.A., GOCAYNE J.D.,
RA  KERLAVAGE A.R., DOUGHERTY B.A., TOMB J.-F., ADAMS M.D., REICH C.I.,
RA  OVERBEK R., KIRKNESS E.F., WEINSTOCK K.G., MERRICK J.M., GLODEK A.,
RA  SCOTT J.L., GEORGHAN N.S.M., WEIDMAN J.F., FUHRMANN J.L., NGUYEN D.,
RA  UTTERBACK T.R., KELLEY J.M., PETERSON J.D., SADOW P.W., HANNA M.C.,
RA  COTTON M.D., ROBERTS K.M., HURST M.A., KAINE B.P., BORODOVSKY M.,
RA  KLENK H.-P., FRASER C.M., SMITH H.O., WOESE C.R., VENTER J.C.;
RT  "Complete genome sequence of the methanogenic archaeon, Methanococcus
RL  jannaschii.";
RL  Science 273:1058-1073(1996).
CC  -1- FUNCTION: CATALYSES A SALVAGE REACTION RESULTING IN THE FORMATION
```

```

CC  OF AMP, THAT IS ENERGICALLY LESS COSTLY THAN DE NOVO SYNTHESIS.
CC  -1- CATALYTIC ACTIVITY: AMP + PYROPHOSPHATE -> ADENINE + 5-PHOSPHO-
CC  ALPHA-D-RIBOSE 1-DIPHOSPHATE.
CC  -1- PATHWAY: PURINE SALVAGE.
CC  -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC  -1- SUBCELLULAR LOCATION: CYTOPLASMIC (POTENTIAL).
CC  -1- SIMILARITY: BELONGS TO THE PURINE/PYRIMIDINE
CC  PHOSPHORIBOSYLTRANSFERASE FAMILY.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL; U67606; AAB99676.1; -.
DR  HSP; P39765; 1A3C.
DR  TIGR; MJ1655; -.
DR  PROSITE; PS00103; PUR_PYR_PR_TRANSFER; 1.
KW  Transferase; Glycosyltransferase; Purine salvage.
SQ  SEQUENCE 183 AA; 20218 MW; E968ECE1 CRC32;

Query Match 67.1%; Score 49; DB 1; Length 183;
Best Local Similarity 75.0%; Pred. No. 9.82e+00;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 76 IPYVIMRK 83
   ||| |||
Qy 1 IPYPIVRK 8

Search completed: Fri Apr 14 23:44:32 2000
Job time : 47 secs.
```

THIS PAGE BLANK (USPTO)





```

RESULT 2
ID 032283 PRELIMINARY; PRT; 52 AA.
AC 032283;
DT 01-JAN-1998 (TEMBLrel. 05, Created)
DT 01-JAN-1998 (TEMBLrel. 05, Last sequence update)
DT 01-NOV-1998 (TEMBLrel. 08, Last annotation update)
DE YZF PROTEIN.
GN YZF.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE; 98044033.
RA KUNST F., OGASAWARA N., MOSER I., ALBERTINI A.M., ALLONI G.,
RA AZEVEDO V., BERTERO M.G., BESSIERES P., BOLOTIN A., BORCHERT S.,
RA BORRIS R., BOURSIER L., BRANS A., BRAUN M., BRIGNELL S.C., BRON S.,
RA BROUILLET S., BRUSCHI C.V., CALDWELL B., CAPUANO V., CARTER N.M.,
RA CHOI S.K., CODANI J.J., CONNERTON I.F., CUMMINGS N.J., DANIEL R.A.,
RA DENIZOT F., DEVINE K.M., DUSTERTHOFT A., EHRLICH S.D., EMERSON P.T.,
RA ENRIAN K.D., ERRINGTON J., FABRET C., FERRARI E., FOULGER D.,
RA FRITZ C., FUJITA M., FUJITA Y., FUNA S., GALIZZI A., GALLERON N.,
RA CHIM S.Y., GLASER P., GOFFEAU A., GOLIGHTLY E.J., GRANDI G.,
RA GUISEPPI G., GUY B.J., HAGA K., HATECH J., HARWOOD C.R., HENAUT A.,
RA HILBERT H., HOLSAPPEL S., HOSONO S., HULLO M.F., ITAYA M., JONES L.,
RA JORIS B., KARAMATA D., KASAHARA Y., KLAER-BLANCHARD M., KLEIN C.,
RA KOBAYASHI Y., KOETTER P., KONINGSTEIN G., KROGH S., KUMANO M.,
RA KURIITA K., LAPIDUS A., LARDINOIS S., LAUBER J., LAZAREVIC V.,
RA LEE S.M., LEVINE A., LIU H., MASUDA S., MAUEL C., MEDIGUE C.,
RA MEDINA N., MELLADO R.P., MIZUNO M., MOESTL D., NAKAI S., NOBACK M.,
RA NOONE D., O'REILLY M., OGAWA K., OGIMURA A., OUDEGA B., PARK S.H.,
RA PARRO V., POHL T.M., PORTELELE D., PORWOLLIK S., PRESCOTT A.M.,
RA PRESCAN E., PUIC P., PURNELLE B., RAPPORT G., REY M., REYNOLDS S.,
RA RIEGER M., RIVOLTA C., ROCHA E., ROCHE B., ROSE M., SADIE I.,
RA SATO T., SCANLAN E., SCHLEICH S., SCHROETER R., SCOFFONE F.,
RA SORIGINI J., SEKOWSKA A., SEROR S.J., SEROR P., SHIN B.S., SOLDI B.,
RA TAKEUCHI M., TAMAKOSHI A., TANAKA T., TERSTRA P., TOGNONI A.,
RA TOSATO V., UCHIYAMA S., VANDENBOL M., VANNIER F., VASSAROTTI A.,
RA VIARI A., WAMBUIT R., WEDLER E., WEDLER H., WEITZENEGGER T.,
RA WINTERS P., WIPAT A., YAMAMOTO H., YAMANE K., YASUMOTO K., YATA K.,
RA YOSHIDA K., YOSHIKAWA H.F., ZUMSTEIN E., YOSHIKAWA H., DANCHIN A.,
RT "The complete genome sequence of the gram-positive bacterium Bacillus
RT subtilis."
RL Nature 390:249-256(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA KUNST F., OGASAWARA N., YOSHIKAWA H., DANCHIN A.,
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; Z99123; CAB15887.1; -.
SQ SEQUENCE 52 AA; 5915 MW; 8AB333D5F CRC32;

Query Match 72.6%; Score 53; DB 2; Length 52;
Best Local Similarity 66.7%; Pred. No. 5.91e+00;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 25 VYPIVRKI 33
:|||||
QY 1 IPYPIVRKL 9

RESULT 3
ID Q9WXW3 PRELIMINARY; PRT; 421 AA.
AC Q9WXW3;
DT 01-NOV-1999 (TEMBLrel. 12, Created)
DT 01-NOV-1999 (TEMBLrel. 12, Last sequence update)
DT 01-NOV-1999 (TEMBLrel. 12, Last annotation update)
DE UDP-N-ACETYLGLUCOSAMINE 1-CARBOXYVINYLTRANSFERASE;
GN TM0108.

Query Match 71.2%; Score 52; DB 2; Length 451;
Best Local Similarity 62.5%; Pred. No. 9.49e+00;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 110 LPYPMYRR 117
:|||||
```

```

OS Thermotoga maritima.
OC Bacteria; Thermotogales; Thermotoga.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 99287316.
RA NELSON K.E., CLAYTON R.A., GILL S.R., GWINN M.L., DODSON R.J.,
RA HAFT D.H., HICKEY E.K., PETERSON J.D., NELSON W.C., KETCHUM K.A.,
RA MCDONALD L., UTTERBACK T.R., MALEK J.A., LINHER K.D., GARRETT M.M.,
RA STEWART A.M., COTTON M.D., PRATT M.S., PHILLIPS C.A., RICHARDSON D.,
RA HEIDELBERG J., SUTTON G.G., FLEISCHMANN R.D., WHITE O., SALZBERG S.L.,
RA SMITH H.O., VENTER J.C., FRASER C.M.;
RA "Evidence for lateral gene transfer between Archaea and bacteria from
RT genome sequence of Thermotoga maritima."
RL Nature 399:323-329(1999).
RN [2]
RP SEQUENCE FROM N.A.
RA NELSON K.E., CLAYTON R.A., GILL S.R., GWINN M.L., DODSON R.J.,
RA HAFT D.H., HICKEY E.K., PETERSON J.D., NELSON W.C., KETCHUM K.A.,
RA MCDONALD L., UTTERBACK T.R., MALEK J.A., LINHER K.D., GARRETT M.M.,
RA STEWART A.M., COTTON M.D., PRATT M.S., PHILLIPS C.A., RICHARDSON D.,
RA HEIDELBERG J., SUTTON G.G., FLEISCHMANN R.D., WHITE O., SALZBERG S.L.,
RA SMITH H.O., VENTER J.C., FRASER C.M.;
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE001697; AAD35202.1; -.
SQ SEQUENCE 421 AA; 45965 MW; A9C30A51 CRC32;

Query Match 71.2%; Score 52; DB 2; Length 421;
Best Local Similarity 55.6%; Pred. No. 9.49e+00;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 83 VPYELVRKM 91
:|||||
QY 1 IPYPIVRKL 9

RESULT 4
ID P74054 PRELIMINARY; PRT; 451 AA.
AC P74054;
DT 01-FEB-1997 (TEMBLrel. 02, Created)
DT 01-FEB-1997 (TEMBLrel. 02, Last sequence update)
DT 01-JAN-1999 (TEMBLrel. 09, Last annotation update)
DE HYPOTHETICAL 50.4 KD PROTEIN.
OS Synchocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synchocystis.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PCC6803;
RA TABATA S.;
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=PCC6803;
RX MEDLINE; 97061201.
RA KANEKO T., SATO S., KOTANI H., TANAKA A., ASAMIZU E., NAKAMURA Y.,
RA MIYAJIMA N., HIROSAWA M., SUGIURA M., SASAMOTO S., KIMURA T.,
RA HOSOUCHI T., MATSUNO A., MURAKI A., NAKAZAKI N., NARUO K., OKUMURA S.,
RA SHIMPO S., TAKEUCHI C., WADA T., WATANABE A., YAMADA M., YASUDA M.,
RA TABATA S.;
RA "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synchocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions."
RL DNA Res. 3:109-136(1996).
DR EMBL; D90911; BAAL8130.1; -.
KW Hypothetical protein.
SQ SEQUENCE 451 AA; 50417 MW; 42DCF091 CRC32;

Query Match 71.2%; Score 52; DB 2; Length 451;
Best Local Similarity 62.5%; Pred. No. 9.49e+00;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 110 LPYPMYRR 117
:|||||
```

QY 1 IPYPIVRK 8

RESULT 5  
ID O64085 PRELIMINARY; PRT; 75 AA.  
AC O64085;  
DT 01-AUG-1998 (TREMBLrel. 07, Created)  
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)  
DT 01-AUG-1998 (TREMBLrel. 07, Last annotation update)  
DE HYPOTHETICAL 9.1 KD PROTEIN.  
GN YOPB.  
OS Bacteriophage SPBc2.  
OC Viruses; dsDNA viruses, no RNA stage; Tailed phages; Siphoviridae.  
RN [1]  
RA LAZAREVIC V., DUESTERHOEF A., SOLDI B., HILBERT H., MAUEL C.,  
RA KARAMATA D.;  
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF020713; AAC13045.1; -;  
KW Hypothetical protein.  
SQ SEQUENCE 75 AA; 9099 MW; 115189E3 CRC32;

Query Match 69.9%; Score 51; DB 9; Length 75;  
Best Local Similarity 85.7%; Pred. No. 1.51e+01;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 28 YPTVRKL 34  
|||  
QY 3 YPIVRKL 9

RESULT 6

ID O31936 PRELIMINARY; PRT; 75 AA.  
AC O31936;  
DT 01-JAN-1998 (TREMBLrel. 05, Created)  
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)  
DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)  
DE YOPB PROTEIN.  
GN YOPB.  
OS Bacillus subtilis.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
OC Bacillus/Staphylococcus group; Bacillus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-168;  
RX MEDLINE; 98044033.  
RA KUNST F., OGASAWARA N., MOSZER I., ALBERTINI A.M., ALLONI G.,  
RA AZEVEDO V., BERTERO M.G., BESIETTES P., BOLOTIN A., BORCHERT S.,  
RA BORRIS R., BOURSTIER L., BRANS A., BRAUN M., BRIGNELL S.C., BRON S.,  
RA BROUILLET S., BRUSCHI C.V., CALDWELL B., CAPUANO V., CARTER N.M.,  
RA CHOI S.K., CODANI J.J., CONNERTON I.F., CUMMINGS N.J., DANIEL R.A.,  
RA DENIZOT F., DEVINE K.M., DUSTERHOFT A., EHRlich S.D., EMMERSON P.T.,  
RA ENTIAN K.D., ERRINGTON J., FABBET C., FERRARI E., FOULGER D.,  
RA FRITZ C., FUJITA M., FUJITA Y., FUMA S., GALIZZI A., GALLERON N.,  
RA GHIM S.Y., GLASER P., GOFFEAU A., COLICHTLY E.J., GRANDI G.,  
RA GUISEPPI G., GUI B.J., HAGA K., HALECH J., HARWOOD C.R., HENAUT A.,  
RA HILBERT H., HOLSAPPEL S., HOSONO S., HULLO M.F., ITAYA M., JONES L.,  
RA JORIS B., KARAMATA D., KASAHARA Y., KLAERR-BLANCHARD M., KLEIN C.,  
RA KOBAYASHI Y., KOETTER P., KONINGSTEIN G., KROGH S., KUMANO M.,  
RA KURITA K., LAPIDUS A., LARDINOIS S., LAUBER J., LAZAREVIC V.,  
RA LEE S.M., LEVINE A., LIU H., MASUDA S., MAUEL C., MEDIGUE C.,  
RA MEDINA N., MELLADO R.P., MIZUNO M., MOESTL D., NAKAI S., NOBACK M.,  
RA NOONE D., O'REILLY M., OGAWA K., OGIMARA A., OUDEGA B., PARK S.H.,  
RA PARRO V., POHL T.M., PORTELELLA D., PORWOLLIK S., PRESCOTT A.M.,  
RA PRESCAN E., PUJIC P., PURNELLE B., RAPAPORT G., REY M., REYNOLDS S.,  
RA RIEGER M., RIVOLTA C., ROCHA E., ROCHE B., ROSE M., SADAIE Y.,  
RA SATO T., SCANLAN E., SCHLEICH S., SCHROETER R., SCOFFONE F.,  
RA SEKIGUCHI J., SEKOWSKA A., SERO S.J., SERROR P., SHIN B.S., SOLDI B.,  
RA SOROKIN A., TACCONI E., TAKAGI T., TAKAHASHI H., TAKEMARU K.,  
RA TAKEUCHI M., TAMAKOSHI A., TANAKA T., TERPSTRA P., TOGNONI A.,  
RA TOSATO V., UCHIYAMA S., VANDENBOL M., VANNIER F., VASSAROTTI A.,  
RA VIARI A., WAMBUTT R., WEDLER E., WEDLER H., WEITZENEGGER T.,  
RA WINTERS P., WIPAT A., YAMAMOTO H., YAMANE K., YASUMOTO K., YATA K.,

RA YOSHIDA K., YOSHIKAWA H.F., ZUMSTEIN E., YOSHIKAWA H., DANCHIN A.;  
RT "The complete genome sequence of the gram-positive bacterium Bacillus  
subtilis";  
RL Nature 390:249-256(1997).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-168;  
RA KUNST F., OGASAWARA N., YOSHIKAWA H., DANCHIN A.;  
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL; Z99115; CAB14013.1; -;  
SQ SEQUENCE 75 AA; 9099 MW; 115189E3 CRC32;

Query Match 69.9%; Score 51; DB 2; Length 75;  
Best Local Similarity 85.7%; Pred. No. 1.51e+01;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 28 YPTVRKL 34  
|||  
QY 3 YPIVRKL 9

RESULT 7

ID O84619 PRELIMINARY; PRT; 112 AA.  
AC O84619;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)  
DE GENOME, PARTIAL SEQUENCE.  
GN A303L.  
OS Paramesidium bursaria chlorella virus 1 (PBCV-1).  
OC Viruses; dsDNA viruses, no RNA stage; Phycodnaviridae; Phycodnavirus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 95133167.  
RA LU Z., LI Y., ZHANG Y., KUTISH G.F., ROCK D.L., VAN ETEN J.L.;  
RT "Analysis of 45 kb of DNA located at the left end of the chlorella  
RT virus PBCV-1 genome";  
RL Virology 206:339-352(1995).  
DR EMBL; U42580; AAC96671.1; -;  
SQ SEQUENCE 112 AA; 13416 MW; 5C07006C CRC32;

Query Match 69.9%; Score 51; DB 14; Length 112;  
Best Local Similarity 44.4%; Pred. No. 1.51e+01;  
Matches 4; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

Db 20 VPFSIIRNL 28  
:|:|:|:|  
QY 1 IYPIVRKL 9

RESULT 8

ID O95214 PRELIMINARY; PRT; 131 AA.  
AC O95214;  
DT 01-MAY-1999 (TREMBLrel. 10, Created)  
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)  
DT 01-MAY-1999 (TREMBLrel. 10, Last annotation update)  
DE BRAIN MYO47 PROTEIN.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Euthera; Primates; Catarrhini; Hominiidae; Homo.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-BRAIN;  
RA MAO Y.M., XIE Y., ZHENG Z.H.;  
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF063605; AAC98697.1; -;  
SQ SEQUENCE 131 AA; 14442 MW; 8C754D73 CRC32;

Query Match 69.9%; Score 51; DB 4; Length 131;  
Best Local Similarity 66.7%; Pred. No. 1.51e+01;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 47 IYPIARRL 55

```
QY 1 IPYPIVRKL 9
|||||
RESULT 9 PRELIMINARY; PRT; 300 AA.
ID Q9WX26
AC Q9WX26;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
DE HYPOTHETICAL 33.1 KD PROTEIN.
GN SCE68.05C.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA MURPHY L., HARRIS D.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA JAMES K.D., PARKHILL J., BARRELL B.G., RAJANDREAM M.A.;
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RX MEDLINE; 97000351.
RA REDENBACH M., KIESER H.M., DENAPATE D., EICHNER A., CULLUM J.,
RA KINASHI H., HOPWOOD D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
DR EMBL; AL079345; CAB45341.1; -.
KW Hypothetical protein.
SQ SEQUENCE 300 AA; 33068 MW; F98702D5 CRC32;

Query Match 69.9%; Score 51; DB 2; Length 300;
Best Local Similarity 75.0%; Pred. No. 1.51e+01;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 176 AYPVREL 183
QY 2 PYPVRL 9
|||||

Query Match 69.9%; Score 51; DB 2; Length 300;
Best Local Similarity 75.0%; Pred. No. 1.51e+01;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 176 AYPVREL 183
QY 2 PYPVRL 9
|||||

RESULT 10 PRELIMINARY; PRT; 520 AA.
ID O66834
AC O66834;
DT 01-AUG-1998 (TREMBlrel. 07, Created)
DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)
DT 01-AUG-1998 (TREMBlrel. 08, Last annotation update)
DE RECOMBINATION PROTEIN RECN.
GN RECN.
OS Aquifex aeolicus.
OC Bacteria; Aquificales; Aquificaceae; Aquifex.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-VF5;
RX MEDLINE; 98196666.
RA DECKERT G., WARREN P.V., GAASTERLAND T., YOUNG W.G., LENOX A.L.,
RA GRAHAM D.E., OVERBEEK R., SNEAD M.A., KELLER M., AUJAY M., HUBER R.,
RA FELDMAN R.A., SHORT J.M., OLSON G.J., SWANSON R.V.;
RT "The complete genome of the hyperthermophilic bacterium Aquifex
RT aeolicus.";
RL Nature 392:353-358(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-VF5;
RA DECKERT G., WARREN P.V., GAASTERLAND T., YOUNG W.G., LENOX A.L.,

Query Match 69.9%; Score 51; DB 2; Length 520;
Best Local Similarity 66.7%; Pred. No. 1.51e+01;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 480 IPYIVREL 488
QY 1 IPYIVREL 9
|||||
RESULT 11 PRELIMINARY; PRT; 565 AA.
ID Q9X210
AC Q9X210;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
DE CONSERVED HYPOTHETICAL PROTEIN.
GN TM1682.
OS Thermotoga maritima.
OC Bacteria; Thermotogales; Thermotoga.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 99287316.
RA NELSON K.E., CLAYTON R.A., GILL S.R., GWINN M.L., DODSON R.J.,
RA HAFT D.H., HICKEY E.K., PETERSON J.D., NELSON W.C., KETCHUM K.A.,
RA MCDONALD L., UTTERBACK T.R., MALEK J.A., LINHER K.D., GARRETT M.M.,
RA STEWART A.M., COTTON M.D., PRATT M.S., PHILLIPS C.A., RICHARDSON D.,
RA HEIDELBERG J., SUTTON G.G., FLEISCHMANN R.D., WHITE O., SALZBERG S.L.,
RA SMITH H.O., VENTER J.C., FRASER C.M.;
RT "Evidence for lateral gene transfer between Archaea and bacteria from
RT genome sequence of Thermotoga maritima.";
RL Nature 399:323-329(1999).
RN [2]
RP SEQUENCE FROM N.A.
RA NELSON K.E., CLAYTON R.A., GILL S.R., GWINN M.L., DODSON R.J.,
RA HAFT D.H., HICKEY E.K., PETERSON J.D., NELSON W.C., KETCHUM K.A.,
RA MCDONALD L., UTTERBACK T.R., MALEK J.A., LINHER K.D., GARRETT M.M.,
RA STEWART A.M., COTTON M.D., PRATT M.S., PHILLIPS C.A., RICHARDSON D.,
RA HEIDELBERG J., SUTTON G.G., FLEISCHMANN R.D., WHITE O., SALZBERG S.L.,
RA SMITH H.O., VENTER J.C., FRASER C.M.;
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE001809; AAD36749.1; -.
SQ SEQUENCE 565 AA; 65087 MW; 77C63CF4 CRC32;

Query Match 69.9%; Score 51; DB 2; Length 565;
Best Local Similarity 55.6%; Pred. No. 1.51e+01;
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 268 ISFPIVRV 276
QY 1 IPYPIVRKL 9
|||||
RESULT 12 PRELIMINARY; PRT; 615 AA.
ID Q9XB2
AC Q9XB2;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
DE SEC D PROTEIN.
GN SEC D.
OS Enterobacter aerogenes (Aerobacter aerogenes).
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Enterobacter.
RN [1]
RP SEQUENCE FROM N.A.
RA KUHN F.C., GIES A.J., SMELTZER M., CRUPPER S.S., SOBIESKI R.J.;
RT "Identification of the secD gene of Enterobacter aerogenes.";
```

RL Submitted (JUN-1999) to the EMBL/GenBank/DBSJ databases.  
DR EMBL; AF163861; AAD44348.1; -  
SQ SEQUENCE 615 AA; 66744 MW; 84E7A77C CRC32;

Query Match 69.9%; Score 51; DB 2; Length 615;  
Best Local Similarity 75.0%; Pred. No. 1.51e+01;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 163 IPYSTVRK 170  
|||: |||  
QY 1 IPYPIVRK 8

RESULT 13  
ID Q51549  
AC Q51549; PRELIMINARY; PRT; 766 AA.  
DT 01-NOV-1996 (TREMBlrel. 01, Created)  
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)  
DE KILLER PROTEIN OF PYOCIN S3.  
GN PYO3A.

OS Pseudomonas aeruginosa.  
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonas group;  
OC Pseudomonas.  
RN [1]

RP SEQUENCE FROM N.A.  
RC STRAIN-P12;  
RX MEDLINE; 95238389.  
RA DUPORT C., BAYSSE C., MICHEL-BRIAND Y.;  
RT "Molecular characterization of pyocin S3, a novel S-type pyocin from  
Pseudomonas aeruginosa."  
RL J. Biol. Chem. 270:8920-8927(1995).  
DR EMBL; X77996; CAA54958.1; -  
SQ SEQUENCE 766 AA; 81434 MW; 66F2A86E CRC32;

Query Match 69.9%; Score 51; DB 2; Length 766;  
Best Local Similarity 66.7%; Pred. No. 1.51e+01;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 679 IPYGEIRKL 687  
|||: |||  
QY 1 IPYPIVRK 9

RESULT 14  
ID Q43048  
AC Q43048; PRELIMINARY; PRT; 834 AA.  
DT 01-JUN-1998 (TREMBlrel. 06, Created)  
DT 01-AUG-1999 (TREMBlrel. 11, Last sequence update)  
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)  
DE PUTATIVE INTEGRAL MEMBRANE GTPASE ACTIVATING PROTEIN, RABGAP DOMAIN  
CONTAINING YEAST MIC1 HOMOLOG.  
GN SPBC215.01 OR SPBC3B9.20.  
OS Schizosaccharomyces pombe (Fission yeast).  
OC Eukaryota; Fungi; Ascomycota; Archiascomycetes;  
OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
OC Schizosaccharomycetes.  
RN [1]

RP SEQUENCE FROM N.A.  
RC STRAIN-972H-;  
RA LYNE M., RAJANDREAM M.A., BARRELL B.G., RIEGER M;  
RL Submitted (NOV-1998) to the EMBL/GenBank/DBSJ databases.  
RN [2]  
RP SEQUENCE OF 1-827 FROM N.A.  
RC STRAIN-972H-;

RA WOOD V., RAJANDREAM M.A., BARRELL B.G., SKELTON J., CHURCHER C.M.;  
RL Submitted (MAR-1997) to the EMBL/GenBank/DBSJ databases.  
DR EMBL; AL033534; CAA22115.1; -  
DR EMBL; AL022070; CAA17800.1; -  
DR PFAM; PF00566; TBC.1.  
KW Hypothetical protein.  
SQ SEQUENCE 834 AA; 95005 MW; C2254AE6 CRC32;

Query Match 69.9%; Score 51; DB 3; Length 834;  
Best Local Similarity 55.6%; Pred. No. 1.51e+01;  
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 82 MPYTIIRKV 90  
|||: |||  
QY 1 IPYPIVRKL 9

RESULT 15  
ID O36900  
AC O36900; PRELIMINARY; PRT; 129 AA.  
DT 01-JAN-1998 (TREMBlrel. 05, Created)  
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)  
DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)  
DE MA-P17 (FRAGMENT).  
GN GAG.

OS Human immunodeficiency virus type 1.  
OC Viruses; Retroviridae; Retroviridae; Lentivirus.  
RN [1]

RP SEQUENCE FROM N.A.  
RX MEDLINE; 97445059.  
RA LEIGH BROWN A.J., LOBIDEL D., WADE C.M., REBUS S., PHILLIPS N.,  
RA BRETTLE R.P., FRANCE A.J., LEEN C.S., MCENAMIN J., MCMILLAN A.,  
RA MAW R.D., MULLCAHY F., ROBERTSON J.R., SANKAR K.N., SCOTT G., WYLD R.,  
RA PEUTHERER J.F.;  
RT "The molecular epidemiology of human immunodeficiency virus type 1 in  
six cities in Britain and Ireland."  
RL Virology 235:166-177(1997).  
DR EMBL; AF014297; AAC58378.1; -  
DR PFAM; PF00540; gag\_p17; 1.  
FT NON\_TER 129  
SQ SEQUENCE 129 AA; 14436 MW; 5201F732 CRC32;

Query Match 68.5%; Score 50; DB 14; Length 129;  
Best Local Similarity 85.7%; Pred. No. 2.40e+01;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 109 YPIVOKL 115  
|||: |||  
QY 3 YPIVRKL 9

Search completed: Fri Apr 14 23:46:34 2000  
Job time : 105 secs.

THIS PAGE BLANK (USPTO)

\*\*\*\*\*

W P E R L H

(TM)

\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsarch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Apr 14 23:48:22 2000; Maspar time 6.10 Seconds  
Tabular output not generated. 34.974 Million cell updates/sec

Title: >US-08-452-843-9  
Description: (1-9) from US08452843.pap  
Perfect Score: 70  
Sequence: 1 IYPPIVRSLS 9

Scoring table: PAM 150  
Gap 15

Searched: 188963 seqs, 23686106 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: a-geneseq36  
1:geneseqp

Statistics: Mean 17.257; Variance 47.133; scale 0.366

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	70	100.0	9	1 R89370	Cw6 consensus peptide	2.60e-01
2	65	92.9	9	1 R89369	Cw6 consensus peptide	1.11e+00
3	57	81.4	9	1 R89371	Cw6 consensus peptide	1.07e+01
4	52	74.3	133	1 W28511	Product of clone L105.	4.23e+01
5	52	74.3	133	1 W50884	Amino acid sequence of	4.23e+01
6	50	71.4	356	1 P70388	D-amino acid oxidase.	7.23e+01
7	50	71.4	356	1 R04066	T-variabilis D-amino a	7.23e+01
8	50	71.4	1676	1 R77604	Pro-C5 polypeptide.	7.23e+01
9	48	68.6	795	1 W97842	Human P2v11 receptor.	1.23e+02
10	47	67.1	216	1 W73419	Human secreted protein	1.60e+02
11	47	67.1	272	1 W56002	Photosynthetic organisi	1.60e+02
12	47	67.1	311	1 W25084	Haemophilus influenzae	1.60e+02
13	47	67.1	319	1 W69735	Human C5a-like protein	1.60e+02
14	47	67.1	319	1 W53896	Human G-protein couple	1.60e+02
15	47	67.1	319	1 W52991	Homoc sapiens clone H96	1.60e+02
16	47	67.1	458	1 W63740	HIV-1 NL-43 gag protei	1.60e+02
17	47	67.1	712	1 Y01772	Brushtail possum zona	1.60e+02
18	47	67.1	1082	1 R65017	PRB2 retinoblastoma tu	1.60e+02
19	47	67.1	1230	1 W17785	Potato tuber solubile s	1.60e+02
20	46	65.7	64	1 Y01423	Secreted protein encod	2.07e+02
21	46	65.7	225	1 W55832	Plasmid pub-B15 human	2.07e+02
22	46	65.7	225	1 W15422	G protein conjugative	2.07e+02
23	46	65.7	343	1 W59907	Human HNHCI32 (G-prote	2.07e+02

24	46	65.7	344	1 W54370	G-protein coupled rece	2.07e+02
25	46	65.7	351	1 W62299	Synechocystis D1 prot	2.07e+02
26	46	65.7	524	1 W35346	Arabidopsis thaliana e	2.07e+02
27	46	65.7	1330	1 R15444	Swine herpes virus-1 m	2.07e+02
28	45	64.3	210	1 R13499	P.denitrificans COB H.	2.68e+02
29	45	64.3	486	1 W37773	Huma glutamine:fructos	2.68e+02
30	45	64.3	713	1 R60101	Canine zona pellucida	2.68e+02
31	45	64.3	716	1 R55200	Feline zona pellucida	2.68e+02
32	45	64.3	716	1 W81810	Feline ZPA protein.	2.68e+02
33	45	64.3	1657	1 W34629	Human C3 protein mutan	2.68e+02
34	45	64.3	1661	1 W34625	Human C3 protein mutan	2.68e+02
35	45	64.3	1663	1 W34606	Wild type human C3 pro	2.68e+02
36	45	64.3	1663	1 W34628	Human C3 protein mutan	2.68e+02
37	45	64.3	1663	1 W34609	Human C3 protein mutan	2.68e+02
38	45	64.3	1663	1 W34817	Human C3 protein mutan	2.68e+02
39	45	64.3	1663	1 W34627	Human C3 protein mutan	2.68e+02
40	45	64.3	1663	1 W34630	Human C3 protein mutan	2.68e+02
41	45	64.3	1663	1 W34615	Human C3 protein mutan	2.68e+02
42	45	64.3	1663	1 W34620	Human C3 protein mutan	2.68e+02
43	45	64.3	1663	1 W40990	Human C3 protein mutan	2.68e+02
44	45	64.3	1663	1 W34607	Human C3 protein mutan	2.68e+02
45	45	64.3	1667	1 W34631	Human C3 protein mutan	2.68e+02

## ALIGNMENTS

## RESULT 1

ID R89370 standard; peptide; 9 AA.

AC R89370;

DT 18-SEP-1996 (first entry)

DE Cw6 consensus peptide derived immunogenic peptide #2.

KW Immunogenic peptide; supermotif; HLA molecule; CTL response;

KW therapeutic; diagnostic; cancer; viral infection; hepatitis B;

KW hepatitis C;

OS Synthetic.

PN W09603140-A1.

PD 08-FEB-1996.

PF 21-JUL-1995; U09234.

PR 21-JUL-1994; US-278634.

PR 23-NOV-1994; US-344824.

PR 30-MAY-1995; US-452843.

PA (CYTE-) CYTEL CORP.

PI Sette A, Sidney J;

PT WPI; 96-116784/12.

PT Compens. comprising immunogenic peptide with supermotif allowing more

than one HLA mol. to bind - used to induce CTL response in patient

and for in vivo and ex vivo therapeutic and diagnostic applications

PS Claim 2; Page 26; 32pp; English.

CC The sequences given in R89362-82 are immunogenic peptides which were

use in the composition of the invention. The composition comprises

an immunogenic peptide of 9-10 residues with a supermotif which

allows binding of more than one HLA molecule. It pref. comprises

two conserved residues, a first at the 2nd position from the N-

terminal is Pro, and a 2nd at the C-terminal is Met. These peptides

are used to induce a CTL response in a patient. They are also

useful in compositions for in vivo and ex vivo therapeutic and

CC diagnostic applications, e.g the treatment of cancer and viral

infections, e.g. hepatitis B and C.

Sequence 9 AA;

Query Match 100.0%; Score 70; DB 1; Length 9;

Best Local Similarity 100.0%; Pred. No. 2.60e-01;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 IYPPIVRSLS 9

|||||

QY 1 IYPPIVRSLS 9

## RESULT 2

ID R89369 standard; peptide; 9 AA.

AC R89369;

DT 18-SEP-1996 (first entry)

```

Query Match      92.9%; Score 65; DB 1; Length 9;
Best Local Similarity 88.9%; Pred. No. 1.11e+00;
Matches      8: Conservative      0; Mismatches 1; Indels 0; Gaps 0;

```

RESULT	3
ID	R89371 standard; peptide; 9 AA.

OS Synthetic.  
PN WO9603140-A1.  
PD 08-FEB-1996.  
PP 21-JUL-1995; U09234.  
PR 21-JUL-1994; US-778634.  
PR 23-NOV-1994; US-344824.  
PR 30-MAY-1995; US-452843.  
PA (CYTE-) CYTEL CORP.  
PI Sette A, Sidney J;  
DRI WPI: 96-116784/12.

WT: 96-116/04/742.

Compsn. comprising immunogenic peptide with supermotif allowing more than one HLA mol. to bind - used to induce CTL response in patient and for in vivo and ex vivo therapeutic and diagnostic applications

Claim 2; Page 26; 32pp; English.

The sequences given in R89363-82 are immunogenic peptides which were used in the composition of the invention. The composition comprises an immunogenic peptide of 9-10 residues with a supermotif which allows binding of more than one HLA molecule. It pret. comprises two conserved residues, a first at the 2nd position from the N-terminal is Pro, and a 2nd at the C-terminal is Met. These peptides are used to induce a CTL response in a patient. They are also useful in compositions for in vivo and ex vivo therapeutic and diagnostic applications, e.g the treatment of cancer and viral infections, e.g. hepatitis B and C.

Sequence 9 AA:

PF	08-AUG-1995:	U12897.
PR	08-AUG-1995:	WO-U12897.
PR	(GEMV )	GENETICS INST INC.
PI	Carlin M, Jacobs K,	Kelleher K, McCoy JM;
PI	WPI: 97-165283/15.	
DR	N-PSDB: T87429.	
DR	Polynucleotide(s)	encoding proteins for treating, preventing and
PT	ameliorating medical conditions - obtained from human activated	
PT	peripheral blood mononuclear cell, and murine adult thymus libraries	
PS	Claim 21; Page 44-45; 61pp; English.	
CC	This sequence was isolated from a murine adult thymus library using	
CC	a trap selecting for nucleotides encoding secreted proteins and	
CC	encodes a protein having homology to various monocytic and other	
CC	chemoattractant proteins.	
SO	Sequence 133 AA:	

```

Db 40 IPYSIVRG 47
    III:III:
Ov 1 IPYPIVRS 8

```

[illegible]

PN	WO9814581-A1.
PD	09-APR-1998.
PE	02-OCT-1997; U17122.
PF	28-AUG-1997; US-058007.
PR	02-OCT-1996; US-027242.
PR	09-OCT-1996; US-028042.
PA	(SCHE ) SCHERING CORP.
PI	Hedrick JA, Zlotnik A;
PI	WFI; 98-240086/21.
PR	N-PSDB; V07113.
PT	Mouse and human CC and CXC chemokine(s) - useful to modulate
PT	physiology or development of cells to treat, e.g. cancerous or
PT	degenerative conditions



PS Claim 1: Pages 78-79; 88pp; English.  
 CC This is the amino acid sequence of the mouse 6CKine (m6CKine) gene, a  
 CC chemokine. It is used in the method of the invention where mouse and  
 CC human CC and CXK chemokines, designated mp4, mCTAP3, m6CKine, h6CKine  
 CC and Chrl9Kine are used to modulate the physiology or the development  
 CC of cells to treat, cancerous or degenerative conditions. The  
 CC chemokines can also be used to generate antibodies, useful in  
 CC immunoassays to measure chemokines, while the nucleic acid sequences  
 CC may be used as components in forensic assays or in situ assays to  
 CC detect chromosomal abnormalities.  
 SQ Sequence 133 AA;

Query Match 74.3%; Score 52; DB 1; Length 133;  
 Best Local Similarity 75.0%; Pred. No. 4.23e+01;  
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 40 IPYSIVRG 47  
 |||||  
 QY 1 IPYPIVRS 8

RESULT 6  
 ID P70388 standard; protein; 356 AA.  
 AC P70388;  
 DT 14-JAN-1991 (first entry)  
 DE D-amino acid oxidase.  
 KW D-amino acid oxidase; Trigonopsis variabilis; cephalosporin;  
 KW oxidative deamination.  
 OS Trigonopsis variabilis.  
 PN J62262994-A.  
 PD 16-NOV-1987.  
 PF 12-MAY-1986; JP-106663.  
 PR (ASAH ) ASAH CHEMICAL IND KK.  
 PA WPI: 87-359677/51.  
 DR N-PSDB; N70609.  
 PT DNA fragment encoding D-amino acid oxidase - which is a useful  
 PT enzyme for the catalytic oxidative deamination of D-amino acids.  
 PS Claim 1; page 583-4; 12pp; Japanese.  
 CC D-amino acid oxidase catalyses the oxidative deamination of D-amino  
 CC acids. It is used in the sepn. of L-amino acids from racemates,  
 CC in the prepn. of ketoic acid from D-amino acid, in amino acid  
 CC analysis, etc. The enzyme can oxidise cephalosporin C to  
 CC 7-beta-(5-carboxy-5-oxopentanamide)cephalosporanic acid, which  
 CC reacts with hydrogen peroxide to give 7-beta-(4-carboxybutanamide)-  
 CC cephalosporanic acid. These cpds. are important intermediates for  
 CC synthesis of cephalosporin type antibiotics.  
 SQ Sequence 356 AA;

Query Match 71.4%; Score 50; DB 1; Length 356;  
 Best Local Similarity 55.6%; Pred. No. 7.23e+01;  
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 66 VSYPIREL 74  
 :|||:|  
 QY 1 IPYPIVRS 9

RESULT 7  
 ID R04066 standard; Protein; 356 AA.  
 AC R04066;  
 DT 03-SEP-1990 (first entry)  
 DE T-variabilis D-amino acid oxidase gene product.  
 KW D-amino acid oxidase; cephalosporin; cephem; E.coli.  
 OS Trigonopsis variabilis.  
 PN EP-364275-A.  
 PD 18-APR-1990.  
 PF 12-OCT-1989; 310483.  
 PR 13-OCT-1988; JP-260332.  
 PA (FUJI) Fufisawa Pharm KK.  
 PI Isogai T, Ono H, Kojo H;  
 DR WPI: 90-11771/16.  
 PT D-amino acid oxidase, prodn. -

PT by culture of E.coli transformants contg. expression vectors  
 PT originated from Fusarium solani M-0718.  
 PS Disclosure; Fig 9; 38pp; English.  
 CC E.coli transformed to express DAO, which catalyses the enzymatic  
 CC conversion of cephalosporin C to 7-beta-(5-carboxy-5-  
 CC oxopentanamide)cephalosporanic acid (keto-7ACA). 7ACA is an  
 CC important starting point for the production of cephem  
 CC antibiotics.  
 SQ Sequence 356 AA;

Query Match 71.4%; Score 50; DB 1; Length 356;  
 Best Local Similarity 55.6%; Pred. No. 7.23e+01;  
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 65 VSYPIREL 73  
 :|||:|  
 QY 1 IPYPIVRS 9

RESULT 8  
 ID R77604 standard; Protein; 1676 AA.  
 AC R77604;  
 DT 15-MAR-1996 (first entry)  
 DE Pro-C5 polypeptide.  
 KW Complement C5; haemolysis; kidney; glomerulonephritis;  
 KW monoclonal antibody; antiinflammatory; antibody engineering;  
 KW humanised antibody.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT Peptide 1..18  
 FT /label= Sig\_peptide  
 FT protein 19..673  
 FT /label= Beta-chain  
 FT cleavage\_site 673..674  
 FT cleavage\_site 677..678  
 FT peptide 674..677  
 FT label= Cleavage\_peptide  
 FT protein 678..1676  
 FT /label= Alpha-chain  
 FT /note= "amino acids 873-892 (854-874 of  
 FT the mature protein) comprise the KSSKS  
 FT epitope"  
 FT peptide 678..751  
 FT /label= C5a  
 FT cleavage\_site 751..752  
 FT /label= Convertase\_cleavage\_site  
 FT modified\_site 911  
 FT /label= N-glycosylation\_site  
 FT modified\_site 1115  
 FT /label= N-glycosylation\_site  
 FT modified\_site 1630  
 FT /label= N-glycosylation\_site  
 PN W09529697-A1.  
 PD 09-NOV-1995.  
 PF 01-MAY-1995; U05688.  
 PR 02-MAY-1994; US-236208.  
 PA (ALEX-) ALEXION PHARM INC.  
 PI Evans MJ, Matis L, Mueller EE, Nye SH, Rollins S;  
 PI Rother RP, Springhorn J P, Squinto SP, Thomas TC;  
 PI Wang Y, Wilkins JA;  
 PI WPI: 95-392923/50.  
 DR Treating glomerulonephritis with antibody against complement C5  
 PT component - to inhibit complement induced cell lysis  
 PS Example 13; Page 82-92; 181pp; English.  
 CC The cDNA sequence of the complement C5 gene transcript predicts a  
 CC secreted pro-C5 precursor of 1676 amino acids (R77604). C5 is a  
 CC beta-globulin heterodimer thought to play a role in the pathogenesis  
 CC of glomerulonephritis (GN). Cleavage of the C5 alpha-chain  
 CC by a convertase enzyme generates anaphylatoxic C5a. Monoclonal  
 CC and humanised recombinant antibodies that recognise the alpha-chain  
 CC KSSKC epitope (R77605) block C5a generation, thereby reducing  
 CC glomerular inflammation and kidney dysfunction associated with GN.  
 SQ Sequence 1676 AA;

Query Match 71.4%; Score 50; DB 1; Length 1676;  
Best Local Similarity 62.5%; Pred. No. 7.23e+01;  
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 829 IPYVVVRG 836

QY 1 IPYPIVRS 8

## RESULT 9

ID W97842 standard; Protein; 795 AA.  
AC W97842;  
DT 07-JUN-1999 (first entry)  
DE Human P2Y11 receptor.  
KW P2Y11; G protein coupled receptor; human; infection; neutropaenia;  
KW agranulocytosis; cancer; leukaemia; diagnosis; therapy.  
OS Homo sapiens.  
FH Key Location/Qualifiers  
FT Modified\_site 474  
FT /note= "putative protein kinase C phosphorylation site"  
FT Modified\_site 600  
FT /note= "N-glycosylation"  
FT Modified\_site 794  
FT /note= "calmodulin-dependent protein kinase phosphorylation site"  
FN W09902675-A1.  
PN 21-JAN-1998.  
PD 09-JUL-1998; BE0108.  
PR 09-JUL-1997; EP-870101.  
PA (EURO-) EUROSREEN SA.  
PI Boeynaems J, Communi D;  
DR WPI: 99-120876/10.  
DR N-PSDB; X07369.  
PT New G protein-coupled receptor - useful for diagnosis, treatment and prevention of neutropaenia, agranulocytosis, infection and cancer  
PS Claim 2; Fig 1; 46pp; English.  
CC This polypeptide comprises a novel human G protein coupled receptor, termed P2Y11 that has selective affinity for ATP. The amino acid sequence was deduced from genomic DNA clones (see X07369). The invention also provides vectors, transformed cells, anti-P2Y11 antibodies, nucleic acid probes, pharmaceutical compositions comprising such products and transgenic animals. Antisense nucleotides (claimed) that hybridise to mRNA are used to decrease activity of P2Y11, while specific antibodies are used to block binding of P2Y11 to its ligand. Probes are used in hybridisation assays to detect expression of P2Y11 at the RNA level, while antibodies are used similarly at the protein level in standard immunoassays, particularly for diagnosis of leukaemia. The transgenic animals are used to determine the effects of varying levels of P2Y11 expression. These animals, and host cells, are used in drug screening methods to identify (ant)agonists that are potentially useful for treatment or prevention of disorders specifically neutropaenia, agranulocytosis, infections and cancer. Host cells are also used to produce recombinant P2Y11.  
SQ Sequence 795 AA;

Query Match 68.6%; Score 48; DB 1; Length 795;  
Best Local Similarity 55.6%; Pred. No. 1.23e+02;  
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 683 VPYHMRVL 691

QY 1 IPYPIVRS 9

## RESULT 10

ID W73419 standard; Protein; 216 AA.  
AC W73419.  
DT 19-FEB-1999 (first entry)  
DE Human secreted protein encoded by Gene No. 23.

KW Secreted protein; human; protein therapy; gene therapy; blood disorder; pathological condition; diagnosis; cancer; neurological disorder; developmental abnormality; foetal deficiency; leukaemia; hepatic disease; immune system disorder; Alzheimer's disease; cognitive disorder; schizophrenia; prostate disease; autoimmune disorder; AIDS.  
OS Homo sapiens.  
FH Key Location/Qualifiers  
FT Misc\_difference 216  
FT /note= "unspecified amino acid"

## PN W09854206-A1.

PD 03-DEC-1998.  
PF 28-MAY-1998; U10868.  
PR 29-AUG-1997; US-056296.  
PR 30-MAY-1997; US-044039.  
PR 30-MAY-1997; US-048093.  
PR 30-MAY-1997; US-048101.  
PR 30-MAY-1997; US-048190.  
PR 30-MAY-1997; US-048356.  
PR 30-MAY-1997; US-050935.  
PR 29-AUG-1997; US-056250.  
PR 29-AUG-1997; US-056293.  
PA (HUMA-) HUMAN GENOME SCI INC.  
PI Carter KC, Dillon PJ, Endress GA, Feng P, Ni J, Rosen CA, Ruben SM, Yu G;  
DR WPI: 99-070209/06.  
DR N-PSDB; Y08833.  
PT New isolated human genes - useful for diagnosis and treatment of, e.g. cancers, neurological disorders, immune diseases, developmental disorders or blood disorders  
PS Claim 11; Page 157-158; 188pp; English.  
CC This sequence is encoded by a cDNA of the invention, designated Gene No. 23. This sequence represents a human secreted protein, and is expressed primarily in immune cells, particularly lymphocytes.  
CC The DNA sequences of the invention and their corresponding secreted polypeptides are useful for preventing, treating or ameliorating medical conditions, e.g. by protein or gene therapy. Also pathological conditions can be diagnosed by determining the amount of the new polypeptides in a sample or by determining the presence of mutations in the DNA sequences. Specific uses are described for each of the DNA sequences and the encoded proteins, based on which tissues they are most highly expressed in, and include developing products for the diagnosis or treatment of cancer, tumours, neurological disorders, developmental abnormalities and foetal deficiencies, blood disorders, leukaemias, diseases of the immune system (including allergies or asthma), hepatic disease, Alzheimer's and cognitive disorders, schizophrenia, prostate diseases, autoimmune disorders and AIDS. The polypeptides are also useful for identifying their binding partners.  
SQ Sequence 216 AA;

Query Match 67.1%; Score 47; DB 1; Length 216;  
Best Local Similarity 71.4%; Pred. No. 1.60e+02;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 140 VPYHIVR 146

QY 1 IPYPIVR 7

## RESULT 11

ID W56002 standard; Protein; 272 AA.  
AC W56002;  
DT 24-JUL-1998 (first entry)  
DE Photosynthetic organism carbonate dehydratase.  
KW Photosynthetic organism; promoter; terminator; carbonate dehydratase; aquatic; blue-green algae; plasmid; growth activity.  
OS Synthetic.  
FH Key Location/Qualifiers  
FT Misc\_difference 4  
FT /note= "encoded by ATC"  
PN J10023891-A.  
PD 27-JAN-1998.  
PF 09-JUL-1996; 179132.  
PR 09-JUL-1996; JP-179132.

PA (SUMO) SUMITOMO CHEM CO LTD.  
PA (CHIK-) ZH CHIKYU KANKYO SANGYO GIJITSU KENKYU.  
DR WPI: 98-152793/14.  
DR N-PSDB: V26252.  
PT New plasmid for improving growth activity, e.g. blue-green algae -  
PT comprises promoter, carbonate dehydratase (sic) and terminator  
PT originating from photosynthetic organism  
PS Claim 5: Page 8-9; 14pp; Japanese.  
CC The present sequence represents a carbonate dehydratase from a  
CC photosynthetic organism, from the present invention. The present  
CC invention describes a new plasmid comprising a promoter, carbonate  
CC dehydratase (sic) and terminator, all of which originate from a  
CC photosynthetic organism (PO) and are capable of functioning in an  
CC aquatic PO. Also described are: (1) a method for the preparation of  
CC the above plasmid, and (2) a microbial host especially an aquatic PO,  
CC e.g. a blue-green algae transformed with the plasmid. The plasmid may  
CC be used to improve the growth activity of an aquatic PO. The transgenic  
CC host may also be used to express the products of the enzyme carbonate  
CC dehydratase (sic).  
SQ Sequence 272 AA;

Query Match 67.1%; Score 47; DB 1: Length 272;  
Best Local Similarity 100.0%; Pred. No. 1.60e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 165 YPIVRS 170

QY 3 YPIVRS 8

RESULT 12  
ID W25084 standard; Protein: 311 AA.  
AC W25084;  
DT 30-DEC-1997 (first entry)  
DE Haemophilus influenzae htrB polypeptide.  
KW Vaccine: htrB gene; Gram-negative bacterium; non-toxic mutant;  
KW pathogen; endotoxin; diagnosis; passive immunisation.  
OS Haemophilus influenzae strain 2019.  
PN W09719688-A1.  
PD 05-JUN-1997.

PF 27-NOV-1996; U18984.  
PR 01-DEC-1995; US-565943.  
PA (AMCY) AMERICAN CYANAMID CO.  
PA (REGC) UNIV CALIFORNIA.  
PA (IOWA) UNIV IOWA RES FOUND.  
PI Apicella MA, Arumugham R, Gibson BW, Lee N, Sunshine MG;  
DR WPI: 97-310355/28.  
DR N-PSDB: T79708.

PT New Gram-negative bacterial pathogen vaccines - comprising a htrB  
PT mutant or an endotoxin isolated from an htrB mutant optionally  
PT conjugated to a carrier protein.  
PS Example 1: Page 61-62; 79pp; English.  
CC This polypeptide comprises the htrB gene product (see also T79708)  
CC of Haemophilus influenzae strain 2019. A claimed vaccine  
CC formulation contains as an active ingredient an htrB mutant of a  
CC Gram-negative bacterial pathogen (GNBP), endotoxin isolated from an  
CC htrB mutant (A) of a GNBP, endotoxin isolated from (A) conjugated  
CC to a carrier protein, or (A) which has been genetically engineered  
CC to express at least one heterologous vaccine antigen, where (A)  
CC lacks one or more secondary acyl chains of lipid A contained in the  
CC GNBP resulting in reduced toxicity when compared to lipid A of the  
CC GNBP. Also claimed is a method for producing endotoxin-specific  
CC antisera for diagnostic assays, or for passive immunisation,  
CC comprising immunising an individual with a vaccine formulation  
CC comprising an active ingredient as above, and collecting antibodies  
CC produced from the immunised individual.  
SQ Sequence 311 AA;

Query Match 67.1%; Score 47; DB 1: Length 311;  
Best Local Similarity 71.4%; Pred. No. 1.60e+02;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 34 LPYPILR 40

QY 1 IPYPIVR 7  
:|||||

RESULT 13  
ID W69735 standard; Protein: 319 AA.  
AC W69735;  
DT 26-OCT-1998 (first entry)  
DE Human C5a-like protein.  
KW Human: C5a-like protein; HCR; diagnosis; complement activation;  
KW inflammation; immunodeficiency; brain de-myelination; neurodegeneration;  
KW allergic reaction; asthma; adult respiratory distress syndrome;  
KW autoimmune disorder; rheumatoid arthritis; systemic lupus erythematosus;  
KW glomerulonephritis; Crohn's disease; cancer; haemodialysis.  
OS Homo sapiens.  
PN W09833908-A1.  
PD 06-AUG-1998.

PF 20-JAN-1998; U01182.  
PR 31-JAN-1997; US-791974.  
PA (INCY) INCYTE PHARM INC.  
PI Bandman O, Coleman R;  
DR WPI: 98-437462/37.  
DR N-PSDB: V50491.

PT Isolated human C5a-like receptor - used to develop products for  
PT diagnosis, prevention and treatment of disorders associated with  
PT complement activation, particularly inflammation  
PS Claim 1: Page 42-43; 59pp; English.  
CC The present sequence represents human C5a-like protein (HCR). The HCR  
CC has similarity to human C5a receptor. Products from the present invention  
CC can be used for the diagnosis, prevention, or treatment of diseases  
CC associated with complement activation. The HCR and agonists can be used  
CC to induce an inflammatory response in a subject who has a diminished  
CC inflammatory response as a result of conditions such as complement  
CC deficiency, immunodeficiency and impaired wound healing. Antagonists or  
CC inhibitors of HCR can be used to prevent inflammation in, e.g. brain  
CC de-myelination and neurodegeneration, allergic reactions, asthma and  
CC adult respiratory distress syndrome, autoimmune disorders such as  
CC rheumatoid arthritis, systemic lupus erythematosus, glomerulonephritis,  
CC and Crohn's disease, post ischaemic myocardial inflammation and necrosis,  
CC skin diseases, septic shock, and inflammatory complications of cancer.  
CC haemodialysis and extracorporeal circulation, infection and trauma. The  
CC products can also be used for detection and drug screening.  
SQ Sequence 319 AA;

Query Match 67.1%; Score 47; DB 1: Length 319;  
Best Local Similarity 71.4%; Pred. No. 1.60e+02;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 237 VPYHIVR 243

QY 1 IPYPIVR 7  
:|||||

RESULT 14  
ID W53896 standard; Protein: 319 AA.  
AC W53896;  
DT 28-AUG-1998 (first entry)  
DE Human G-protein coupled receptor HLYAZ61.  
KW HLYAZ61; G-protein coupled receptor; human; therapy;  
KW diagnosis; infection; HIV-1; HIV-2; pain; cancer; anorexia;  
KW bullimia; asthma; Parkinson's disease; acute heart failure;  
KW atherosclerosis; hypotension; hypertension; urinary retention;  
KW osteoporosis; angina pectoris; myocardial infarction; ulcer;  
KW allergy; benign prostatic hypertrophy; neurological disorder;  
KW psychosis; anxiety; schizophrenia; manic depression; delirium;  
KW dementia; mental retardation; dyskinesia; Huntington's disease;  
KW Gilles de la Tourette's syndrome.  
OS Homo sapiens.  
PN EP-837128-A2.  
PD 22-APR-1998.

PF 16-OCT-1997; 308207.  
PR 21-OCT-1996; US-734349.  
PA (SMIK) SMITHKLINE BEECHAM CORP.

Search completed: Fri Apr 14 23:49:05 2000  
Job time : 43 secs.

PI Bergsma DJ, Ellis CE;  
DR WPI; 98-219111/20.  
DR N-PSDB; V23658.  
PT DNA encoding G-protein coupled receptor protein - useful for  
PT producing recombinant peptides and in gene therapy  
PS Claim 13; Fig 1a-c; 38pp; English.  
CC This polypeptide comprises HLYA261, a novel human G-protein coupled  
CC receptor containing 7 hydrophobic regions that may represent  
CC membrane spanning domains. Its amino acid sequence was deduced from  
CC a cDNA clone (see v23658) isolated from a human leukocyte cDNA  
CC library. This polynucleotide can be utilised in the recombinant  
CC production of HLYA261 in host cells. HLYA261 polypeptides may be  
CC employed for therapeutic purposes, including treatment of bacterial,  
CC fungal, protozoan and viral infections, particularly infections  
CC caused by HIV-1 and HIV-2, pain, cancers, anorexia, bulimia,  
CC asthma, Parkinson's disease, acute heart failure, atherosclerosis,  
CC hypotension, hypertension, urinary retention, osteoporosis, angina  
CC pectoris, myocardial infarction, ulcers, allergies, benign  
CC prostatic hypertrophy and psychotic and neurological disorders  
CC including anxiety, schizophrenia, manic depression, delirium,  
CC dementia or severe mental retardation, and dyskinesias, such as  
CC Huntington's disease or Gilles de la Tourette's syndrome. The  
CC polypeptide can also be used in a claimed method for identifying  
CC compounds which bind to and activate or inhibit a receptor for  
CC HLYA261. Also disclosed are diagnostic assays for detecting  
CC diseases related to altered concentrations of HLYA261 polypeptides.  
SQ Sequence 319 AA;

Query Match 67.1%; Score 47; DB 1; Length 319;  
Best Local Similarity 71.4%; Pred. NO. 1.60e+02;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 237 VPYHIVR 243  
QY 1 IPYPIVR 7

RESULT 15  
ID W52991 standard; Protein; 319 AA.  
AC W52991;  
DT 03-AUG-1998 (first entry)  
DE Homo sapiens clone H963\_20 protein.  
KW secreted protein; human; PBMC; peripheral blood mononuclear cells.  
OS Homo sapiens.  
FH Key Location/Qualifiers  
FT Peptide 1..35 /note= "signal sequence"  
FT  
FN W09807859-A2.  
PD 26-FEB-1998.  
PF 22-AUG-1997; U14874.  
PR 23-AUG-1996; US-702344.  
PA (GEMV ) Genetics Inst Inc.  
PI Marberg F, McCoy JM, Lavallie ER, Racie LA, Treacy M, Spaulding V,  
PI Jacobs K;  
DR WPI; 98-169163/15.  
DR N-PSDB; V21240.  
PT New nucleic acid encoding secreted proteins from human cells -  
PT useful e.g. as immuno-modulators, antitumour agents, promoters of  
PT tissue growth, haemostatic and thrombolytic agents etc.  
PS Claim 34; Pages 63-64; 79pp; English.  
CC The sequence is that encoded by the clone H963\_20 which was  
CC isolated from a human adult PBMC cDNA library using methods  
CC selective for cDNAs that encode secreted proteins.  
SQ Sequence 319 AA;

Query Match 67.1%; Score 47; DB 1; Length 319;  
Best Local Similarity 71.4%; Pred. NO. 1.60e+02;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 237 VPYHIVR 243  
QY 1 IPYPIVR 7